Engineering Letters, 14:2, EL_14_2_2 (Advance online publication: 16 May 2007)

Wearable Real-Time Stereo Vision for the Visually Impaired

G. Balakrishnan, G. Sainarayanan, R. Nagarajan and Sazali Yaacob

***Abstract*— Visually impaired find their navigation difficult as they often lack the needed information for bypassing obstacles and hazards. Electronic Travel Aids (ETAs) are devices that use sensor technology to assist and improve the blind user's mobility in terms of safety and speed. Modern ETAs does not provide distance information directly and clearly. This paper proposes a method for determining distance using a stereo matching method to help blind individuals for their navigation. The system developed in this work, named Stereo Vision based Electronic Travel Aid (SVETA), consists of a computing device, stereo cameras and stereo earphones, all molded in a helmet. An improved area based stereo matching is performed over the transformed images to calculate dense disparity image. Low texture filter and left/right consistency check are carried out to remove the noises and to highlight the obstacles. A sonification procedure is proposed to map the disparity image to stereo musical sound, which has information about the features of the scene in front of the user. The sound is conveyed to the blind user through stereo headphones. Experimentations have been conducted and preliminary investigations have proven the viability of this method for applying in real time environment.**

***Index Terms*— Stereo matching, Electronic Travel Aid, Disparity, Stereo Vision, Sonification.**

I. Introduction

Most aspects of the dissemination of information to aid navigation and cues for active mobility are passed to human through the most complex sensory system, the vision system. This visual information forms the basis for most navigational tasks and so with impaired vision an individual is at a disadvantage because appropriate information about the environment is not available. According to World Health Organization census, around 180 million people worldwide are visually disabled, of those 40 to 45 million populations are totally blind [1]. This population is expected to double by the year 2020.

Two low technology aids for the blind, the long cane and the guide dog [2], have been used by blind for many years. A number of electronic mobility aids using sonar [3]–[4] have also been developed to detect obstacles, but market acceptance is rather low as useful information obtainable from them are not significantly more than that from the long cane. The outputs produced are also complex for user understanding. Recent research efforts are being directed to produce new navigational system in which digital video camera is used as vision sensor. In The vOICe [5], the image is captured using single video camera mounted on a headgear and the captured image is scanned from left to right direction for sound generation. The top portion of the image is converted into high frequency tones and the bottom portion into low frequency tones. The loudness of sound depends on the brightness of the pixel. Similar work has been carried out in NAVI [6] where the captured image is resized to 32 X 32 and the gray scale of the image is reduced to 4 levels. With the help of image processing technique the image is differentiated into objects and background. The objects are assigned with high intensity values and the background is suppressed to low intensity values. Here the processed image is converted into stereo sound where the amplitude of the sound is directly proportional to intensity of image pixels, and the frequency of sound is inversely proportional to vertical orientation of pixels. By using single camera the distance information cannot be obtained effectively.

The distance is one of the important aspects for collision free navigation for blinds. In order to incorporate the distance information, stereo cameras have to be used. The manner in which human beings use their two eyes to see and perceive the three-dimensional world has inspired the use of two cameras to model the world in three dimensions. The different perspectives of the same view seen by two cameras lead to a relative displacement of the same objects or the same points in world reference (called disparity). The size and direction of these disparities can be utilized for depth estimation. The depth of a point is inversely proportional to the amount of disparity.

Using stereo vision for blind navigation application is in early stages and only limited research efforts has been reported in it. In Optophone [7], to obtain a depth map an edge detection routine is applied to images from two cameras. Disparity is calculated using the edge features of both the images. The depth map is then converted into sound using the method

Manuscript received May 31, 2006. Authors wish to thank Ministry of Science, Technology and Innovation, Malaysia for funding the research through Universiti Malaysia Sabah under IRPA code: 03-02-10-0043/EA0041.

G. Balakrishnan is with the School of engineering and Information Technology, University Malaysia Sabah, 88999, Kotakinabalu, Malaysia. (phone: +6088-320000-3558, fax: +6088-320348, e-mail: g_krishbalu@yahoo.com).

G. Sainarayanan is with the School of engineering and Information Technology, University Malaysia Sabah, 88999, Kotakinabalu, Malaysia. (email: jgksai@ums.edu.my).

R. Nagarajan is with the School of Mechatronics, Northern Malaysia University College of Engineering, 01000 Kangar, Perlis, Malaysia. (email: nagarajanums@yahoo.com).

Sazali Yaacob is with the School of Mechatronics, Northern Malaysia University College of Engineering, 01000 Kangar, Perlis, Malaysia. (e-mail: sazali@kukum.edu.my).

applied in The vOICe system [5] where, the top portion of the image is converted into high frequency tones and the bottom portion into low frequency tones. The loudness of sound is directly proportional to intensity of the pixel. In Optophone the disparity map of all the edge features in the images is obtained. The user will find difficult to locate the object since unwanted edge features will also exist. With only the edge information, it will be difficult to identify the object.

Another pioneering work by Zelek et. al. involves stereo camera and was designed to provide information about the environment through tactile feedback to the blind [8]. The system comprises of a laptop, a stereo head with two cameras and a virtual touch tactile system. The tactile system is made up of piezo-electric buzzers attached to each finger on a glove worn by the user. Here the cameras capture images, and the disparity is calculated from those images. The depth information are conveyed to the user by stimulating the fingers. In this work no image processing efforts are undertaken to highlight the object information in the output. More over the system suffers in stereo matching.

Another important work reported in this area is the visual support system developed by Yoshihiro Kawai and Fumiaki Tomita [9]. The prototype system has a computer, stereo camera system with three small cameras, headset with a microphone and headphone and sound space processor. The images captured by small stereo cameras are analysed to obtain 3D structure, and object recognition is performed. The results are then converted to user via 3D virtual sound. The prototype developed is huge and not portable. It can be applicable only in indoor environment. From the literature, it is clear that efforts are made to use stereo vision in Electronic Travel Aid. But the recent researches have faced the problems in stereo matching, information transference and in making the system portable. There are no commercial stereo vision based electronic travel aids so far. In this paper, methods have been developed to overcome the problems encountered in earlier researches. An improved area based stereo matching is employed in this paper to calculate the distance information and obstacle information is conveyed to the blind using the musical tone concept.

II. Overview of SVETA System

The prototype system is named as Stereo Vision based Electronic Travel Aid (SVETA). The hardwares used in this work are small enough to be carried out easily. The SVETA system consists of a headgear molded with stereo cameras and stereo earphones. The Compact Computing Device (CCD) is placed in a specially designed pouch. The user has to wear the pouch, wherever he goes with the SVETA system. The stereo camera selected is a compact, low-power digital stereo head with an IEEE 1394 digital interface. It consists of two 1.3 mega pixel, progressive scan CMOS imagers mounted in a rigid body, and a 1394 peripheral interface module, joined in an integral unit. The CCD is handy with a high performance 500MHz Intel mobile Celeron processor with 256 MB RAM. The helmet can be worn over the head. The SVETA prototype system is shown in Figure 1(a).

The stereo cameras are placed in the front of the headgear, located slightly above the position of the eyes as shown in Figure 1(b). The stereo cameras capture the visual information infront of the blind user. The captured images are then processed using the proposed methodology in the CCD. The information about the obstacle is conveyed to the blind user by musical tones and voice commands through stereo earphones.

(a)

(b)

Figure 1: SVETA (a) Prototype system (b) Blind user wearing SVETA

III. Stereo Vision

Stereovision is a paradigm to calculate the distance of an object by analyzing the two images of an object acquired from two different directions or orientations. Image acquisition, camera modeling, feature acquisition, image matching, depth determination are the various steps in this paradigm [10]. Image matching (stereo matching) is an important and difficult step in this paradigm. The stereo matching algorithms available in the current literature are broadly classified into two classes: area-based and feature-based algorithms [11]. Feature-based algorithms need preprocessing of stereo images to find the positions of features such as edges, corner points and line segments. These techniques provide sparse disparity, i.e. only at the positions of features. The features in one image may be occluded in the other image. Hence, these techniques are less preferred. Area-based (window-based) algorithms are advantageous, as they provide dense disparity. These algorithms perform matching at each pixel, using absolute intensity value of pixels. In this paper, an improved area based stereo matching algorithm has been proposed for depth determination in SVETA.

IV. STEREO MATCHING

The image size used in this work is 240 x 320. The Stereo matching method proposed in this work undergoes the following steps:

A. Preprocessing

The standard form for stereo processing assumes that the two images are from pinhole cameras of the same focal length, and are co-planar, with scan lines and focal centers aligned horizontally. In practical, the stereo images acquired from the cameras are distorted and focal centers are not aligned horizontally. In this step the distortions in the images are rectified and converted into standard images. In order to minimize the search space to one dimensional, calibration is performed. There are two parts in calibration: internal calibration, dealing with the properties of the individual cameras and especially lens distortion and external calibration, the spatial relationship of the cameras to each other. Both internal and external calibrations are performed using the method proposed by Zhengyou Zhang [12]. From the internal and external parameters, the calibration procedure computes an image warp for rectifying the left and right images. In stereo rectification, the images are effectively rotated about their centers of projection to establish the ideal stereo setup: two cameras with parallel optical axes and horizontal epipolar lines. Having the epipolar lines horizontal is crucial for correspondence finding in stereo, as stereo looks for matches along horizontal scanlines.

Figure 2 shows a pair of original images (I_O) acquired from the cameras. In the original images (I_O), effect of lens distortion can be noticed at the corners of the images. Also, the images are not aligned vertically. Figure 3 is the result of calibrating the stereo cameras and then rectifying the two original images. Now the rectified images (I_R) are aligned vertically, and all scene lines are straight in the images.

B. Image Transforms

Area correlation compares small patches among images using correlation. Blind navigation application involves real time stereo images. In environment like outdoor scenes, there are possibilities of uneven lighting illumination for left and right stereo images due to different focus of cameras or due to position of light source. Correlation of image areas may be affected by illumination differences among stereo images. Stereo matching by correlating the raw intensities will produce poor results. This paper proposes two improvements to tackle the problems of stereo matching.

1. Laplacian of Gaussian transforms for better matching, and
2. Low Texture Filter that invalidates uncertain matches at low texture region.

In this work, the stereo matching attempts to compensate the matching by correlating the transform of the raw intensity images. Laplacian of Gaussian (LoG) transform with standard deviation of 0.5 is used for image transform. Laplacian of an image highlights regions of rapid intensity change. It is applied to the images that have first been smoothed with Gaussian filter in order to reduce its sensitivity to noise. The LoG function is given by

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}\left[1-\frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

where
x,y is the pixel coordinates,
σ is Gaussian standard deviation

Figure 2: Original left and right stereo images (I_O)

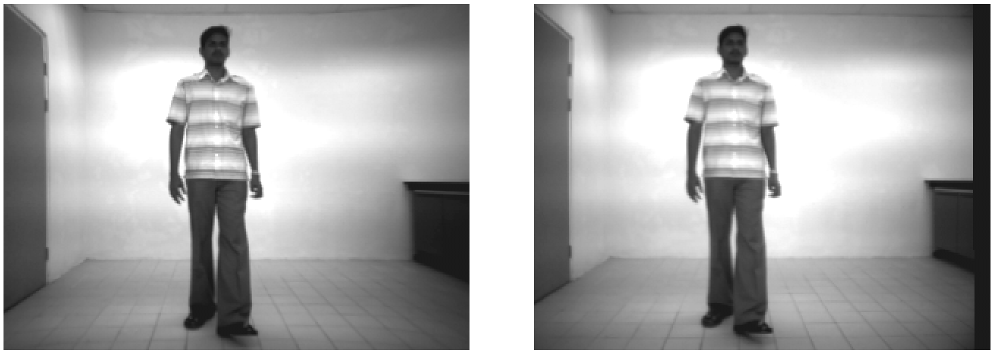


Figure 3: Restored left and right calibrated images (I_R)

After Laplacian of Gaussian transform, the area where rapid intensity change occurs will be highlighted. Let the stereo image pairs after Laplacian of Gaussian transform be I_L. Stereo matching of Laplacian of Gaussian transformed images (I_L) usually derive dense disparity image with some false matches. Figure 4 shows the disparity image obtained after stereo matching of stereo images I_L. It is clear from the Figure 4, there are some false matches and errors at depth discontinuities. The errors due to false matches can be removed using Low Texture filter and the depth discontinuity errors can be minimized using Left/Right consistency check.

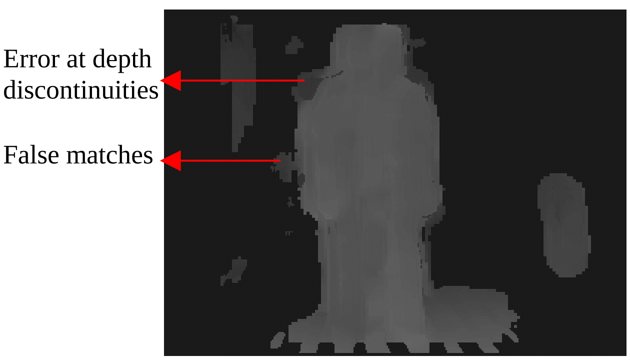


Figure 4: Disparity image obtained from images I_L

C. Low Texture Filter

Low texture filter locates uniformly scattered features over the stereo images I_L, having the property that the corresponding points are unambiguously findable in subsequent images. Therefore, low texture filter minimize the probability that important obstacles will be missed, and also attempts to choose areas which can be found easily by correlator. The low texture filter is defined as follows

The directional variances in brightness for each pixel along each of the four directions are computed as

$$V_1 = \sum_{(x,y)\in S} [I(x,y) - I(x,y+1)]^2 \qquad (2)$$

$$V_2 = \sum_{(x,y)\in S} [I(x,y) - I(x+1,y)]^2 \qquad (3)$$

$$V_3 = \sum_{(x,y)\in S} [I(x,y) - I(x+1,y+1)]^2 \qquad (4)$$

$$V_4 = \sum_{(x,y)\in S} [I(x,y) - I(x+1,y-1)]^2 \qquad (5)$$

where I is the image, S is the local window and (x,y) is the pixel coordinates. If an area has no variance along any of the four directions, it is difficult to match with its counterpart. The reason is that there may be many identical areas along that direction, and it will be difficult to decide which of them forms actual match. The minimum value of the directional variances is taken as the interest operator for any central pixel (Xc, Yc).

Therefore the interest operator is given as

$$I(Xc, Yc) = \min(V_1, V_2, V_3, V_4) \qquad (6)$$

The local maxima of the interest value that exceed a certain threshold are determined. These points are selected for matching. Since the variance measure depends on adjacent pixel differences, it responds to high frequency noise in the image. In order to avoid this undesirable effect, Low texture filter is applied to Laplacian of Gaussian transformed images (IL) where high frequency noises are minimized.

Given a feature in left image, stereo matching is performed to find the corresponding region in the right image. The correlation measure selected in this work is Sum of Absolute Difference (SAD) because of its computational efficiency. The sum of absolute difference measure is defined by

$$\sum_{x,y} \left| I_1(x,y) - I_2(x+d,y) \right| \qquad (7)$$

where
I_1 is the window in left image,
I_2 is the window in right image and
d is the disparity range.
The window size used in this work was 11 x 11. The size of the window was selected based on experimentation. The effect of stereo matching based on different window sizes is explained in the following section.

The low texture filter gives high confidence to areas that are textured in intensity, since non-texture areas are subject to ambiguous matches. Therefore, using low texture filter, featureless area and simple edges are avoided. The errors at depth discontinuity can be removed using left/right consistency check. The left/right check looks for consistency in matching from a fixed left image region to a set of right image regions, and back again from the matched right region to a set of left regions. It is particularly useful at range discontinuities, where directional matching will yield different results.

Figure 5 shows the final disparity image obtained after applying the low texture filter and left/right consistency check. Areas with insufficient texture are rejected as low confidence and they appear black in the picture. Left/Right check operator removes the errors at the portions of the image with disparity discontinuities. The combination of low texture filter and left/right check has proven to be the most effective at eliminating bad matches. The total computation time for stereo matching is approximately 1 second.

Figure 5: Disparity image after post filtering

D. Effect of Window Size and Non-Calibration

The size of the correlation window determines the amount of pixels used for correlation. The effect of noise is reduced by increasing the number of pixels and thus the size of the correlation window. However, bigger correlation windows are more likely to cover areas where depth varies. Usually, rectangular windows are used for the sake of computational performance [13]. Windows of variable size can be implemented but the computation time involved is not appropriate for this real time application [14]. Hence upon experimentations, the window size of 11 x 11 seems to provide better results compared to other size.

As described earlier, the standard form for stereo processing assumes that the scan lines of the two images are aligned horizontally. If all the above described methods are applied to the stereo uncalibrated images, vague disparity map is obtained as shown in Figure 6. The difference between the disparity maps obtained from the calibrated and uncalibrated is visually obvious, and it is found that image measures correlate well with calibrated images. If the input images are not calibrated, there will be few valid stereo matches, and the left/right check will reject many of these.

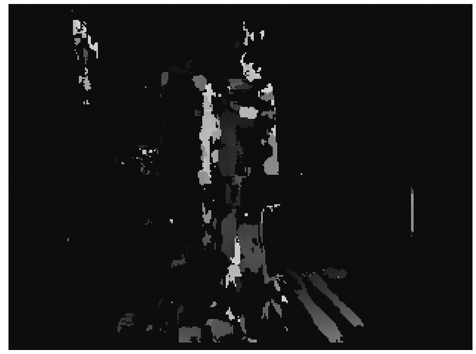


Figure 6: Disparity image obtained from uncalibrated stereo images

Figures 7 – 10 shows disparity image of a person approaching towards the stereo camera, obtained using the proposed method where the disparity variation is clearly visible. Higher disparities (closer objects) are indicated by bright color and low disparity objects (lying far away) are represented by varied gray shades based on its value. From the Figures 7 – 10, as the person approach close to the stereo camera the disparity image changes from dark color to bright color indicating its increase in value. The following section describes in detail about the mapping of disparity image into stereo musical sound.

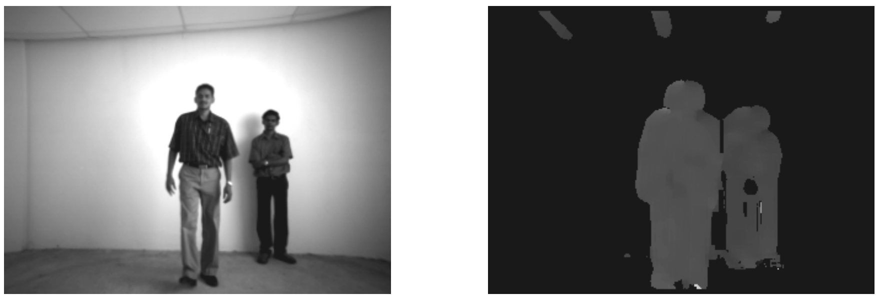
V. SONIFICATION

The final disparity image is then converted to stereo musical sound using the following sonification methodology. The effective audible frequency range extends from about 20 Hz to around ten thousand hertz, although it depends entirely on the individual. The audible range is divided into octaves. An octave is a frequency range from a frequency f1 to f2 such that f2 is twice that of f1 in terms of cycles or hertz. The human ear is logarithmic and is sensitive to frequency octaves. The audible frequency is then comprised of many octaves. Even a frequency range from 20 Hz to 40 Hz is defined as an octave [15].

In most of the western musical instruments the frequencies are arranged in such a manner that they are in a geometric series. That is, the frequency deviation between any key and the key immediately to its left is a constant, the constant being equal to the twelfth root of two or 1.059. Even though there is a degree of freedom for selecting the range of an octave (whether it is from 240 to 480 Hz or 254 to 508 Hz etc.), the western music defines a standard octave called the Middle A octave starting from 440 Hz.

(a) (b)

Figure 7: (a) Stereo left image (b) Disparity image



(a) (b)

Figure 8: (a) Stereo left image (b) Disparity image

(a) (b)

Figure 9: (a) Stereo left image (b) Disparity image

(a) (b)

Figure 10: (a) Stereo left image (b) Disparity image

With the help of the above rules, a set of musical tones can be incorporated for image sonification. For fast computation, the disparity image is resized to 32 x 32 pixels. Through a set of experiments, it is found that the octave frequency of 440 Hz to 880 Hz produces pleasing tones. With this octave, 12 musical notes are developed. Let f(1), f(2),..., f(12) be these 12 octave frequencies. Then the music pattern can be generated by

$$M(j) = \sin\left(2\pi\, f(j)\ t\right),\ \text{j=1,2,…,12} \tag{8}$$

where M(j) is the musical note generated for f(j)th frequency and t varies from 0 to a desired total duration of the acoustic information presented to the blind.

Different musical tones can be generated by a combination of these notes. In this work, three notes are combined to form the required set of musical tones. Four half steps between first and second note and three half steps between second and third note define the major chords. Here, eight tones including some major chords are generated using these notes. Every preceding four rows are grouped and assigned with one musical tone. These musical tones are assigned in such a way that high frequency tones are generated for the top portion of the image and low frequency tones are assigned to the lower portion of the image. So, each pixel in an image is assigned with a sample of musical tone based on its position in the image.

The conversion of image into sound involves taking one column of image pixels at a time starting from left most column and generating sound pattern in succession. The sound pattern generated is hence given by

$$S(j) = \sum_{i=1}^{32} I(i, j) M(i, j) \tag{9}$$

where S(j) is the sound pattern produced from column j of the image, j = 1,2,…,16 and j = 32,31,…17 for stereo type scanning, I(i,j) is the intensity value of (i,j)th element, and M(i,j) is the sample of musical tone for (i,j)th pixel.

The sound pattern from each column is appended to construct the sound for the entire image. The scanning of the image is performed in such a way that a stereo sound is produced. The scanning is performed from leftmost column towards the centre and from right most column towards the centre, simultaneously. In this stereo type scanning, the sound patterns created from the left part of the image is given to the left earphone and the sound patterns of right part to the right earphone simultaneously. Different tones are produced for different shapes, orientation and intensities. Hence the sound pattern generated by this sonification method is able to differentiate objects based on its position, shape and distance. In this method, if any obstacles approach very close to the user, at an approximate distance of 165 cm, the disparity value will be very high and a voice command is triggered in order to alert the user about the nearing obstacle. Also if the user approaches any non-texture region like wall or door, no disparity information will be obtained. Hence an audible voice is alarmed to the user to be aware of forthcoming obstacle. The most advantage of this method is that since musical tones are used, the sound generated will be pleasing to the user and continuous use will not fashion loss of interest. Figure 11 shows the 3D visualization of musical sound generated for the disparity image in Figure 9, through the proposed method.

VI. Testing

The total computation time of the SVETA system to process the image and to feedback the information is about 1.25 seconds. For a blind user to navigate freely, the information about the obstacles such as shape, size, position and distance are to be known. If all those information are conveyed to the blind user, then he can move autonomously and collision free among the obstacles. These features are very important for blind navigation. In SVETA, the distance is mapped to the amplitude of the sound, the shape and size of objects are represented with different combinations of musical tones, vertical orientation is mapped to frequency of the sound and horizontal orientation is represented using left and right channel of the sound. The proposed sonification method is thus compared and tested for their capabilities. In this experimental study, the blind and non blind people were tested to find obstacles based on its distance and its characteristics by using the developed sonification method. The main purpose was to determine whether the subjects can identify the obstacles easily. The experiment was tested using several groups, with 12 people in each group. The data such as object position, size and distance from each method were collected from each subject. Also, every subject is inquired regarding pleasantness of sound and their response is monitored.

Simulated test images were initially used. They were of white, light gray or dark gray shapes of squares and circles with a black background. White color corresponds to the object in close distance. Different gray shades are assigned to objects based on their corresponding distance. Subjects were also tested with objects of different shapes and sizes placed at different positions, a number of times in different days. The test results obtained are tabulated as in Table 1.

Table 1: Test results for identifying the characteristics of obstacles

Object Characteristics	Percentage predicted
	Musical Octave Method

Position (Top/Bottom)	78%
Position (Left/Right)	100%
Shape	88%
Size	91%
Distance	98%
Pleasantness	91%

The developed prototype for vision substitution scheme was tested on visually handicapped volunteers. A blind person was trained with some basic shapes of images and he was tested to identify the uniqueness and pleasantness of the sound. It was found that he was able to recognize the basic shapes like square, circle, triangle, rectangle and diagonal bars separately and combined. He was also asked to recognize objects in the indoor environment both stationary and moving. With repeated training, he was able to walk successfully in indoor without collision and also along corridor without hitting on walls. Suggestions from the blind volunteer regarding pleasantness and discrimination of sound patterns were also considered in frequency scaling. The magnitude of sound is adjusted to an optimal level so that the blind can easily hear sounds coming even form outside of the system. Work is progressing to train the blind person in identifying the real time outdoor scene in front of him through the musical sound pattern produced by this prototype. In this research, it was observed that the blind user identified voice commands as useful and interesting feature for their navigation. Figure 12 shows the blind volunteer being tested in indoor environment.

Figure 12: Blind user tested with SVETA in indoor environment

VII. CONCLUSION

A portable, wearable stereo vision system (SVETA) has been developed for the purpose of enhancing mobility and navigation of a visually impaired person. An improved area based stereo matching and musical octave based image sonification algorithm for real time blind navigation application were proposed in this paper. The proposed stereo matching method is suitable for real time distance computation. The sound produced using musical octave concept is more pleasant for continuous hearing. The system has been tested with several users and the feedback proves the applicability of the newly developed system. The system suffers some matching errors in bright outdoor environment. In future, the work is extended towards using the system in all environments, improving the computation time and in developing a user interface which allows task setting by the user by voice.

REFERENCES

[1] World Health Organization, Webpage: http://www.who.int//mipfiles/2400/AllenFoster.pdf

[2] F. Wong, R. Nagarajan, S. Yaacob and A. Chekima, "Electronic travel aids for visually impaired – A guided tour", Proceedings of Conference in Engineering in Sarawak, 2000, pp. 377-382.

[3] L. Kay, "An Ultrasonic sensing probe as a mobility aid for the blind", Ultrasonics 1964, 2: 53-59.

[4] N. Pressey, "Mowat Sensor", Focus 3, 1977, pp. 35-39.

[5] P. Meijer, "An Experimental System for Auditory Image Representations", IEEE Transactions on Biomedical Engineering, Feb 1991, 39(2):112-121.

[6] G. Sainarayanan, "On Intelligent Image Processing Methodologies Applied to Navigation Assistance for Visually Impaired", Ph. D. Thesis, University Malaysia Sabah, 2002.

[7] M. Capp and P. Picton, "The Optophone: an electronic blind aid", Engineering Science and education Journal, 2000, 9(3): 137-143.

[8] J. Zelek, S. Bromley, D. Aamar and D. Thompson, "A haptic glove as a tactile vision sensory substitution for way finding", Journal of Visual Impairment and Blindness, 2003, pp. 621–632.

[9] Y. Kawai and F. Tomita, "A support system for visually impaired persons to understand three-dimensional visual information using acoustic interface", IEEE 16th conference on Pattern Recognition, 2002, 3: 974 – 977.

[10] Z. Brown and D. Hager, "Advances in computational stereo", IEEE Trans. Pattern Analysis and Machine Intelligence, Aug 2003, 25(8):993-1008.

[11] U. R. Dhond, J. K. Aggarwal, "Structure from stereo", IEEE Transaction on systems, man and cybernatics, 1989, 19: 1489 - 1510.

[12] "Camera Calibration", webpage: http://www.vision.caltech.edu/bouguetj/calib_doc/

[13] H. Hirschmüller, P. R. Innocent and J. M. Garibaldi, "Real Time Correlation based Stereo Vision with Reduced Border Errors", International Journal of Computer Vision , 2002, 47(1-3): 229-246.

[14] A. Fusiello, V. Roberto, and E. Trucco, "Efficient stereo with multiple windowing", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 858-863.

[15] R. Elliott, Fundamentals of Music, New Jersey : Prentice-Hall, 1971.

G. Balakrishnan is currently a research officer in electrical and electronics engineering in Universiti Malaysia Sabah. He received his B.E., M.E., and Ph.D. degrees, from Baradidhasan University, India, Bharathiar University, India, and University Malaysia Sabah, Malaysia respectively in 2001, 2003 and 2006. His research interests are in the areas of vision rehabilitation, image processing and machine vision.

G. Sainarayanan is currently a Lecturer in electrical and electronics engineering in Universiti Malaysia Sabah, Malaysia. He received his B.E., M.E., and Ph.D. degrees, respectively, from Annamali University, India, Bharathiar University, India, and University Malaysia Sabah, Malaysia, in 1998, 2000, and 2002. His research interests are in the areas of vision rehabilitation, medical imaging, and intelligent control. He has published more than 45 papers in refereed journals, book chapters, conferences, and research reports.

R. Nagarajan received his B.E. degree (Hons) in electrical and electronics engineering from Madras University in 1961, M. Tech. degree in control

engineering from IIT Kanpur in 1969, and Ph.D. degree in adaptive control from Madras University in 1977. Currently he is a Professor in the School of Mechatronics, Northern University College of Engineering, Malaysia. He was a Visiting Professor at Osaka Prefecture University, Osaka, under the JSPS program from October to November 2004. His current research interests include robot control, stereo image processing, and intelligent process control. He has received several awards for his research contributions which include medals from International Invention and Exhibition in Malaysia and Geneva. He is a Life Fellow of Institution Engineers, India, a Senior Member in IEEE, USA, and a Member of UICEE, Australia.

Sazali Yaacob was born in 1960.He received his Bachelor of Engineering degree in electrical engineering in 1986 from University Malaya. He received his Master of Science degree in system engineering from the University of Surrey, U.K., and later his Doctor of Philosophy degree in control engineering from the University of Sheffield, U.K., in 1987 and 1995, respectively. He was promoted to Associate Professor in 1998 at University Malaysia Sabah, Malaysia, and was consequently appointed as the First Dean of the School of Engineering and Information Technology until 2004. Now, he is working at the Northern University College of Engineering, Malaysia, as Professor in the Center of Engineering Study in Mechatronics. He has publications in the national and international journals and conference proceedings. He has supervised a number of postgraduate students in either Master or Doctor of Philosophy levels. His research interests are in artificial intelligence applications such as in the fields of acoustics, modeling, and control. He was granted a Charted Engineer status by Engineering Council, United Kingdom, in 2004.