ON STEREO PROCESSING PROCEDURE APPLIED TOWARDS BLIND NAVIGATION AID - SVETA

G.Balakrishnan[1] *G.Sainarayanan*[1] *R.Nagarajan*[1] *Sazali Yaccob*[2]
[1]AI Research Group, School of Engineering and Information Technology
Universiti Malaysia Sabah, Locked Bag No. 2073, 88999, Malaysia
[2] Northern Malaysia University College of Engineering, Perlis, Malaysia
[1] g_krishbalu@yahoo.com

ABSTRACT

This paper presents a portable traveling support system using stereo image processing for the visually impaired people's navigation. This system named as SVETA consists of a helmet molded with stereo cameras in the front, wearable computer over the top and stereo earphones. The two cameras capture the visual information infront of the blind user. The captured images are then processed using the proposed methodology in wearable computer. The methodology includes stereo image processing module to calculate distance through disparity. The information is conveyed to the blind through the set of earphones in terms of musical tones. Voices are also used to inform blind user when the obstacle is very close to him. Experimentations were conducted in both indoor and outdoor environment, and the results of this experiment verified the practicability of the newly developed system.

1. 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 [10]. This population is expected to double by the year 2020.

Two low technology aids for the blind, the long cane and the guide dog, have been used by blind for many years. A number of electronic mobility aids using sonar [6, 8] 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 [7, 9]. The distance is one of the important aspects for collision free navigation for blinds. By using single camera the distance information cannot be obtained effectively. 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.

Only limited research has been done in blind navigation using stereo cameras. In Optophone [2], 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 applied in The vOICe system [7]. Here 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 pioneered work by Zelek et.al. involves stereo camera and was designed to provide information about the environment through tactile feedback to the blind [12]. 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 is conveyed to the user by stimulating the fingers. In this work no image processing method is applied and the information about the object is not direct. More over the system suffers in stereo matching.

One more important work reported in this area is the visual support system developed by Yoshihiro Kawai and Fumiaki Tomita [5]. 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

0-7803-9243-4/05/$20.00 ©2005 IEEE

converted to user via 3D virtual sound. The prototype developed is huge and not portable. It can be applicable only in indoor environment. In this paper, methods to estimate distance information from stereo images and conveying those information through sound have been proposed.

2. OVERVIEW OF SYSTEM

The main objective of this work is to develop a portable and wearable system for blind to navigate through the obstacles by knowing its distance information. The hardwares used in this work are small enough to be carried out easily. The SVETA (Stereo Vision based Electronic Travel Aid) system consists of a helmet moulded with stereo cameras, wearable computer and stereo earphones. The stereo camera 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 processing computer is of wearable type with 500MHz processor and 256 MB RAM. The helmet can be worn over the head. The stereo camera is placed in the front and the wearable computer is placed over the top of the helmet as shown in Fig. 1(b). Fig. 1(a) shows a volunteer wearing the SVETA system.

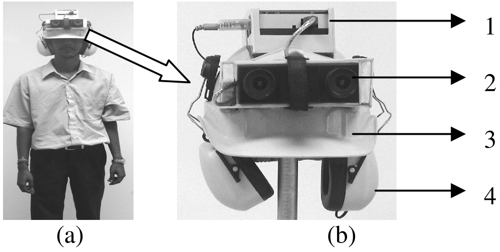

Fig. 1. (a) Person wearing SVETA system
(b) The SVETA system. 1: Wearable computer, 2: Stereo camera, 3: Helmet and 4: Stereo earphone

3. STEREO IMAGE PROCESSING

Stereo analysis is the process of measuring depth of an object based on a comparison of the object projection on two or more images [1]. The major steps involved in depth calculation are preprocessing, establishing correspondence and post filtering to reduce mismatch errors. Stereo correspondence is the fundamental problem to be solved. Correspondence techniques can be classified into two broad categories namely Feature based technique and Area based technique. In feature-based techniques, depth estimation is performed only for certain points of interest, such as corners or edges. These techniques are faster since only a small subset of image points are used, however the generated disparity map may be less accurate due to some of the erroneously excluded important features. In area based techniques, displacement information are calculated for the entire image. This provides enhanced depth information, however estimation errors may occur especially in low-texture regions.

3.1. 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, it is difficult to achieve parallel scan lines. In order to minimize the search space to one dimensional, calibration is performed. In the proposed method the stereo images acquired from stereo cameras are calibrated using the method proposed in [3]. Fig. 2 shows the original stereo pair acquired from the cameras. In Fig. 2 the corners of images are little bit curved. This is due to lens distortion. Also the images are not aligned horizontally. Fig. 3 shows the results after rectifying the original images. Now the images are aligned horizontally and the lens distortion is removed.

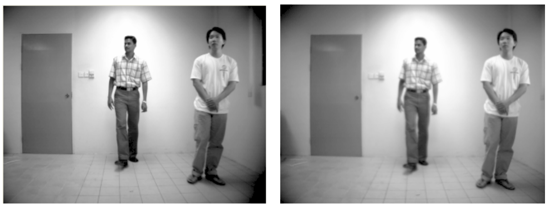

Fig. 2. Original left and right stereo images

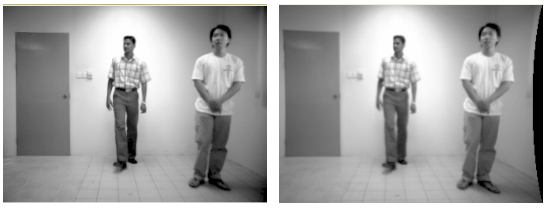

Fig. 3. Restored left and right calibrated image

3.2. Stereo Correspondence

In blind navigation application, dense disparity information are needed. So, in the proposed work area or correlation based stereo matching is performed. Correlation of image areas is disturbed by illumination, perspective, and imaging differences among images. Area correlation methods usually attempt to compensate by correlating not the raw intensity images, but some transform of the intensities. There are three basic types of transforms: 1. *Normalized intensities:* Each of the intensities in a correlated area is normalized by the average intensity in the area. 2. *Laplacian of Gaussian (LOG)*: The laplacian measures directed edge intensities over some area smoothed by the Gaussian and 3. *Nonparametric:* These transforms are an attempt to deal with the problem of outliers, which tend to overwhelm the correlation measure, especially using a square difference.

The algorithm implemented in this work makes use of LOG transform with standard deviation of 1.0 using absolute difference correlation. The window size selected in this work is 11 x 11. The LOG transform and absolute difference norm were chosen because they give good quality results, and can be optimized on standard instruction sets available on DSPs and microprocessors [11]. Fig. 4(a) shows the disparity image obtained after performing area correlation over stereo pairs shown in Fig. 3. The higher disparities (close objects) are represented by brighter white color. If the object is very close, then it is represented by white color. If the object is far away (low disparity), it is represented by dark gray color.

3.3. Post Filtering

Dense disparity images usually contain false matches which must be filtered. In this work, interest operator and left/right check are performed for post filtering the disparity image. An interest operator gives high confidence to areas that are textured in intensity, since flat areas are subject to ambiguous matches. 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. Fig. 4(b) shows the post filtered disparity image obtained from original disparity image shown in Fig. 4(a). Areas with insufficient texture are rejected as low confidence and hence they appear black in the figure. In practice, the combination of an interest operator and left/right check has proven to be the most effective at eliminating bad matches [11].

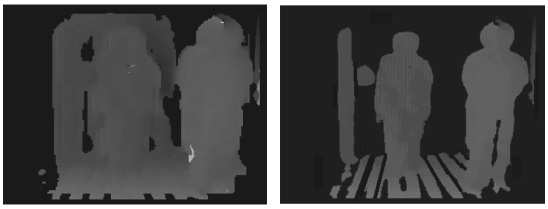

Fig. 4. (a) Original disparity image **(b)** Disparity image after post filtering

Some of the results obtained from the proposed stereo matching methods are shown in Fig. 5 and 6.

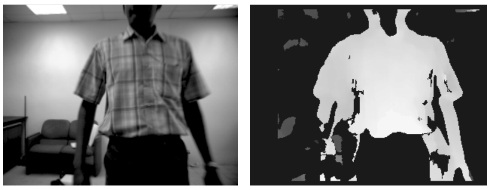

Fig. 5. Stereo left Image and its disparity image

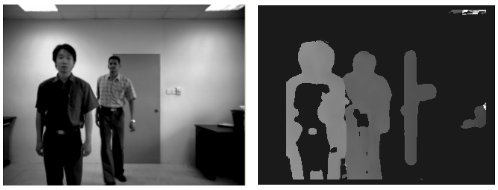

Fig. 6. Stereo left Image and its disparity image

4. 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 [4]. 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.

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

$$M(j) = \sin\left(2\pi f(j)\ t\right), \quad j=1,2,\ldots,12 \qquad (1)$$

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

$$S(j) = \sum_{i=1}^{32} I(i, j) M(i, j) \qquad (2)$$

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. 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. The scanning is performed from leftmost column towards the centre and from right most column towards the centre, simultaneously. Different tones are produced for different shapes. Hence the sound pattern generated by this sonification method is able to differentiate objects based on its position, shape and distance. Also in this method, if any obstacles approach close to the user, at an approximate distance of 165 cm, the disparity will be very high and a voice command of stop will be played in order to alert the user about the nearing 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.

5. TESTING

The total computation time of the SVETA system to process the image and to feedback the information is about 1.25 seconds. The SVETA system has been tested in indoor environment and in some of the outdoor environments with blind and non-blind people. The volume of the sound increases with the increase in disparity (ie with the decrease in distance). The users feel trouble free in predicting the distance and also in locating the obstacle. But some of the users find difficulties in guessing the object characteristics. Also the stereo matching method provides vague results in very bright sunny environments. Blind users are being trained currently to adapt with the SVETA system.

6. CONCLUSION

A portable, wearable stereo vision system (SVETA) has been developed for the purpose of enhancing mobility and navigation of a visually impaired person. Distance is an important aspect for blind navigation. Stereo cameras are used to calculate the distance. Area based stereo matching is performed over rectified stereo images in order to obtain dense disparity map. The noises in the disparity image are removed up using interest operator and left/right consistence check. The most interesting aspect in this work is the musical tone auditory representation. The system has been tested with several users and the feedback proves the applicability of the newly developed system. Some matching errors occur 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 voice.

7. REFERENCES

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

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

[3] “Camera Calibration”, webpage: http://www.vision.caltech.edu/bouguetj/calib_doc/

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

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

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

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

[8] Pressey N, “Mowat Sensor”, Focus 3, pages 35-39, 1977.

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

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

[11] Yang R and Pollefeys M, “Multi-resolution real-time stereo on commodity graphics hardware”, *International Conference on Computer Vision and Pattern Recognition,* pages 211-220, 2003.

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