Stereo Camera based Artificial Vision for Blind through Hearing

G.Balakrishnan[+] G.Sainarayanan[+] R. Nagarajan[+] Sazali Yaacob* M. Rizon*
[+]AI Research Group, School of Engineering and Information Technology
Universiti Malaysia Sabah, Locked Bag No. 2073, 88999, Malaysia
**School of Mechatronic Engineering*
Northern Malaysia University College of Engineering
01000 Kangar, Perlis, Malaysia
E-mail: rizon@kukum.edu.my

Abstract

In this paper, a design of real time methodology for object identification and image sonification applied to vision substitution for the blind is presented. The proposed methodology includes intelligent image processing, stereopsis and image sonification modules. The hardware system consists of stereo digital video cameras as vision sensors, stereo headphone and a laptop computer. The images captured from stereo cameras are processed to obtain isolated object image. The stereopsis module performs area based stereo correspondence over the object image and calculates the disparity. The object is given a preference based on its distance and position using fuzzy rule base for collision free navigation by blind user. Finally the information is conveyed to the blind through stereo sound in the form of musical tones. In the proposed algorithm distance information of the object is determined with less computation compared to conventional area based stereo correspondence method. Experiments were conducted in an indoor environment, and the results obtained are promising.

Keywords:

Vision substitution, Fuzzy inference system, Stereopsis, Blind navigation

1 Introduction

Human navigation consists of two distinct components: sensing the immediate environment for impediments to travel (e.g., obstacles and hazards) and navigating to remote destinations beyond the immediately perceptible environment. Informations about way finding are passed to humans through the most sophisticated sensory system, the vision system. The visually impaired are at a considerable disadvantage, as they often lack the needed information for bypassing obstacles and hazards. According to World Health Organization census, around 180 million people worldwide are visually disabled, of those 40 to 45 million are totally blind [7]. This population is expected to double by the year 2020.

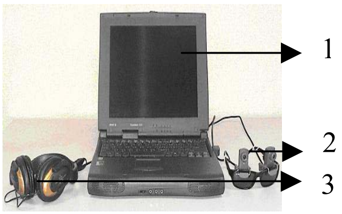
During the past three decades, several researchers have introduced devices called *Electronic Travel Aids* (ETAs) [6], that use sensor technology to assist and improve the blind users mobility in terms of safety and speed. In early type of ETAs, the users had to scan the environment constantly and continuously. Locating an obstacle in the path was time consuming and similar to having a long cane. But in recent years, due to advancement in the development of high-speed computers and sensory devices, attempts to design sophisticated equipment for the vision substitution are in progress.

The concept of using video camera as vision sensor had been introduced in Peter Meijer's portable system, the vOICe [4]. The image captured is scanned from left to right direction for sonification or 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. In Peter Meijer's work, image-processing efforts to enhance the object properties are not undertaken. The sound produced from the unprocessed image will contain more information of the background rather than object. This may be the reason for the blind user having difficulties to distinguish the object and the background through sound from earphones.

Navigational Assistance for Visually Impaired (NAVI) [5] makes use of single camera and the captured image is resized to 32 X 32 and the gray scale of the image is reduced to 4 levels. The image is differentiated into objects and background. The objects are assigned high intensity and the background is suppressed. 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. Both in vOICe and NAVI the distance between the user and the obstacle cannot be obtained directly by the users. The distance is one of the important aspects for collision free navigation for blinds. In order to incorporate the distance information, stereo cameras are used in this work. 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.

2 Experimental Setup

A prototype system is designed for this work. The hardware model constructed for this vision substitution system has a sunglass fitted with two mini video cameras, stereo earphone and a laptop computer. The two cameras are displaced from each other by a distance of 5 cm. Both the cameras are adjusted to same focus by experimentations. Figure 1 shows the experimental setup used in this work. When the sunglass is worned by the blind user, the stereo cameras capture the scene infront of the user.



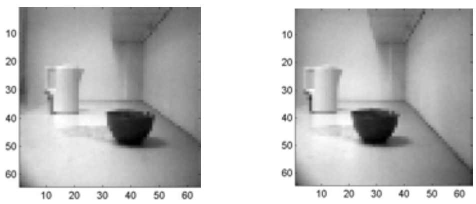
(a)

1 – Laptop Computer
2 – Sunglass fitted with stereo camera
3 – Stereo headphone

Figure 1: Prototype system

3 Image Processing Methodology

In this paper, objects or obstacles are identified and are assigned preference based on their position and distance. The first step in this methodology is the image acquisition, in which scene in front of blind is captured using both cameras simultaneously. The image captured from camera is a color image of size 352 x 288. Processing the image with original size will increase computation time. This work involves real time computation. So the computation time is critical and has to be minimized. Therefore pre-processing is undertaken to reduce the computation time, where both the left and right stereo images are converted to gray scale intensity image and resized to 64 x 64. Figure 2 shows the preprocessed left and right image. The main task involved here is to identify and assign preference to the objects. In this work, the objects in both the left and right camera images are identified by locating its edges.



(a) (b)

Figure 2: Resized Gray Scale Intensity Image (a) left image (b) right image

3.1 Image Post Processing

Edge detection is one of the important human vision properties as it has the ability to recognize the object boundaries. The goal of edge extraction is to provide useful structural information about object boundaries. Canny edge detector is used for edge detection because this method uses two thresholds to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. Canny edge detector is an optimum edge detector [2]. Figure 3 shows the edge features of the right camera image extracted using canny edge detector. Similar processing is undertaken for left camera image also.

Figure 3: Edge Image of Right Stereo Image

In this work a region enclosed within edges is considered as an object. It is realized that edge detection alone is not enough to extract the closed boundary of an object from the image. Therefore further processing is undertaken to connect the edges to form a meaningful object. Edge linking process is required to assemble these edges into meaningful edges. Morphological operation using dilation [1] is proposed to link the broken edges. The region within the closed boundary is considered as an object. The intensity of pixels in the region is enhanced to higher level using flood fill operation. Each object in the image is labeled. It is found that some of the extraneous edges still exist after flood fill operation. Hence noise removal operation is performed to remove extraneous edges present in the image without eliminating the desired objects. Erosion and dilation operations are undertaken to smooth the objects [1]. Thus image with only objects will be obtained. Figure 4 shows the image obtained after flood fill and noise removal method.

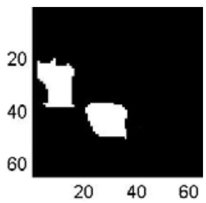


Figure 4: Right Image with only objects

3.2 Image Object Isolation

Binary images of both left and right camera image with only objects are derived from the previous section. This binary image is mapped with the resized gray scale intensity image and new gray scale intensity image with only objects is derived. These processes are preformed for both left and right camera images to obtain two gray scale images with only objects. Figure 5 shows the left and right images with only objects.

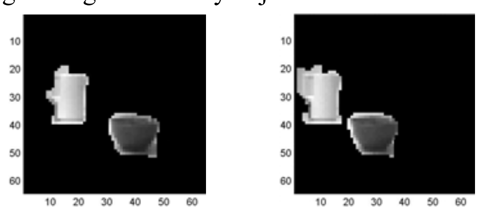


Figure 5: Grayscale Image with Only Objects of Left and Right Camera Images

4 Stereo Vision Methodology

After obtaining the isolated object image, the disparity has to be determined in order to calculate the distance between the blind user and the obstacle. The concept of stereopsis is employed here. Given two camera images, if it is possible to identify the image locations that correspond to the same physical point in space from both the camera images, then it is possible to determine its three-dimensional location. In order to calculate the disparity correspondence or matching of two images has to be done. In this work, area based correspondence method is applied over two isolated object images. To decrease the total computation time, the maximum disparity is bounded to a certain range and thus avoids examining an entire row for each pixel. Since only the isolated object images are used for correspondence, the mismatch error is less and also the computation time is reduced when compared to conventional area based and feature based techniques.

From the disparity map obtained, it is observed that in some areas, the disparity value within the object varies due to some mismatch. For uniform disparity value in an object, histogram is used. In a histogram, disparity value that occurs most within the object is found and that particular value is assigned to all pixels within the object Figure 6 shows the disparity map obtained after assigning uniform value.

Figure 6: Disparity Map

5 Object Preference

A Real time environment contains more than one object. If all objects are given same preference, then the blind user finds it difficult to identify those objects, which are obstacles in their path for navigation. So each object has to be assigned some priority for collision free navigation. In this work objects are given preference based on two important conditions. One is the distance of the object from the user and another one is the position of the object. In human vision system, the eyes mostly concentrate on a particular object rather than the background, which gets less focus. The object of interest will be usually at the center of the sight. In this paper, the center of the image is considered as the center of sight. Any object that is located in the central region with high disparity will be considered for high preference and objects located outside the central region with low disparity will be less preferred.

In order to check whether the object is located in the central region, object characteristics such as size and centroid have to be determined. Based on these characteristics, the ratio of the object area lying within the central region can be calculated. But with these inputs, devising an algorithm to compute object preference is very difficult. In order to overcome this uncertainty fuzzy logic is applied. Four main characteristics are measured for object preference assignment. They are size of an object, Euclidean distance between the object centroid and image centroid, ratio of objects lying within the center of an image and the distance through disparity of an object [1]. These characteristics are applied as input to fuzzy logic algorithm. Each characteristic is expressed using three membership functions namely low, medium and high. Membership functions are expressed using a trapezoidal curve. The output is object preference, which has four trapezoid curve memberships. The defuzzification is performed using centroid method. 81 rules are formed based on four inputs.

From the fuzzy output the objects are given different intensity values based on their preference. The object with very high preference is symbolized by white intensity and the object with least preference is symbolized with very dark gray intensity. The object preference based on fuzzy rule based system is shown in Figure 7.

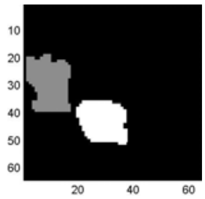


Figure 7: Object Preference Based on Fuzzy Output

The fuzzy output image is then converted to stereo sound using the following sonification methodology.

6 Image Sonification

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 really 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. The f1 can be any number but f2 is double the f1. Even a frequency range from 20 Hz to 40 Hz is defined as an octave [3].

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 concept, musical tones can be incorporated for image sonification. By experimentation it is found that octave frequency of 440 Hz to 880 Hz produce pleasing music. In the developed method, this octave frequency of 440 Hz to 880 Hz is selected. With this octave, 12 musical notes are developed. Let f (1,..,12) be the 12 octave frequencies.

The music pattern generated is given by

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

where

M(j) is the musical note generated for f(j) frequency and 't' varies from zero to desired total duration of the acoustic information.

Different musical tones are generated with the combination of these notes. In this work three notes are combined to form musical tones. Four half steps between first and second note and three half steps between second and third note define major chords. Here eight tones including some major chords are generated using these notes. The image to be sonified is resized to 32 X 32 for reducing the computation time. Musical tones are assigned in such a way that high frequency tones occupies the top portion of the image and low frequency tones are assigned to lower portion of the image. So each pixel in an image is assigned with samples of musical tones based on their position in the image.

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

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

where S(i) is the sound pattern for column i of the image I(i,j) is the intensity value of (i,j)th element, M(i,j) is the samples 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 picture is performed in such a way that stereo sound is produced. In this stereo type scanning, the sound patterns created from the left half side of the image is given to left earphone and sound patterns of right half side to right earphone simultaneously. The scanning is performed from leftmost column towards the centre and from right most column towards the centre.

Since the sound produced are differentiated into left and part, objects that lie to the left of the user will produce sound only in the left earphone and the object that lies to the right will be heard only in the right earphone. 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. 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. Prototype has been designed to minimize the hardware and using the system in outdoor environment also. Figure 8 shows the design of future prototype.

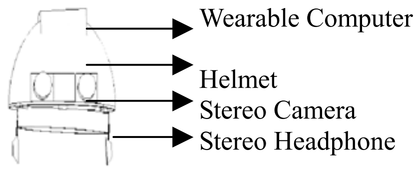


Figure 8: Future Wearable Prototype

7 Conclusions

Distance information is one of the most important criteria for blind navigation. In this work stereo cameras are used to find the distance information and the information is conveyed to the blind user through musical tones. Experiments were conducted in an indoor environment, and the results obtained are promising. The computation time of this method is also fast compared to conventional area based methods. There are some errors due to stereo mismatching. In the near future, the work will be continued towards achieving higher accuracy, minimizing the hardwares and testing the system with the blind people.

Acknowledgment: 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.

Reference

[1] Balakrishnan G, Sainarayanan G, Nagarajan R, Yaacob S, "Blind Navigation Using Stereopsis With Fuzzy Rule Base", *International Conference on Artificial Intelligence in Engineering and Technology*, Malaysia, pp. 232-237, 2004.

[2] Canny J, "A Computational Approach to Edge Detection", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 6, pp. 721-741, 1984.

[3] Elliott R, "Fundamentals of music", New Jersey: Prentice-Hall.

[4] Meijer P, "An Experimental System for Auditory Image Representations", *IEEE Transactions on Biomedical Engineering*, Vol 39, pp 112-121, 1991.

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

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

[7] World Health Organization, Internet site address: http://www.who.int//mipfiles/2400/AllenFoster.pdf