Object Identification and Colour Recognition for Human Blind

R. Nagarajan, G. Sainarayanan, Sazali Yaacob and Rosalyn R Porle[1]
Artificial Intelligence Research Group, School of Engineering and Information Technology,
Universiti Malaysia Sabah, Locked Bag No. 2073, 88999 Kota Kinabalu, Sabah, Malaysia
[1]email: rlyn39@yahoo.com

Abstract

Navigation Assistance for Visually Impaired (NAVI) is a vision substitute system designed to assist blind people for autonomous navigation. NAVI working concept is based on 'image to sound' conversion. The vision sensor, which is mounted on a sunglass, captures the image infront of blind user. NAVI processing unit processes the captured image and enhances the significant vision data by employing a set of image processing procedures. The processed information is then presented as a structured form of acoustic signal and it is conveyed to the blind user using a set of earphones. This paper proposes a new object identification and colour recognition module for NAVI system. Image processing techniques are proposed to identify objects in the captured image and then the processed image is transformed into stereo sound patterns. Colour information from the interested object is evaluated to determine the colour of the object. The colour output is informed to the blind user through 'verbal voice'.

1. Introduction

World Health Organization (WHO) reported that there are currently about 45 million blind people worldwide and every year this number is increases by 1 to 2 million [16]. Blind people's navigation is restricted because they do not receive enough information about the objects or obstacles in their environment. Electronic Travel Aids (ETAs) are electronic devices designed to improve autonomous navigation of blind people [15]. ETAs design varies from the sizes, the type of sensor used in the system, the method of conveying information and also the method of usage. In 1992, Peter Meijer has introduced 'image to sound' concept in his invention, *The vOICe* [5]. Image of the scene infront of blind user is captured using video camera and it is transformed into sound pattern. The intensity of the pixel of the image is transformed into loudness. In Peter Meijer's work, image processing efforts to enhance the object properties are not undertaken. In real world scene, background fills more area than the objects. It is also noted that most of the background is of light intensity and the sound produced on it will be of high amplitude compared to the objects in the scene. As a result, the sound produced from the unprocessed image may contain more information of the background rather than object. This may be the reason for the blind user's difficulties to distinguish the object and the background of the environment [6].

Navigation Assistance for Visually Impaired (NAVI) is a vision substitute system designed to aid blind people for autonomous navigation. NAVI working concept is also based on 'image to sound' conversion whereby a captured image is processed and transform to stereo sound pattern. The stereo sound pattern reflects the nature of scene and of any objects nearby or in the front of blind user. NAVI system is intended to be applicable in indoor environment. The image processing methodology in NAVI is designed as a model of human vision system that facilitates more importance to the properties of objects rather than background of the scene. The captured image from the vision sensor is resized and quantitized into four grey levels. Objects are identified with its grey value. The identified objects are then enhanced and the background is suppressed using a clustering algorithm [7]. The processed image is mapped to stereo sound patterns and the amplitude of the sound depends on the grey intensity of that particular object.

Due to lighting effect, single object may have more than one grey level value. With the former consideration of objects, there will be high possibilities that objects in the image were not highlighted or background was highlighted. Noise elimination procedure was not undertaken in the early method and thus noise still exists in the image. Furthermore, grey level quantitization in the earlier stages of processing would result reduction of information in the image. This information is imperative for object identification. Realizing this, an improved image processing for NAVI is proposed in this paper, which emphasizes on a new object identification methodology. Objects in colour image are extracted using its closed boundary. The proposed procedure also aims to eliminate noise in the image. The processed image is then transformed into stereo sound pattern.

From the sound production, the blind user can discern object's properties such as size, shape and orientation of objects. Colour of object can also be utilized for object identification since most objects can be described by its colour. From the experience with blind people during training, it is noted that the blind people are interested in knowing the colour of the object. Most blind people are not blind at birth. So they have familiarity toward colour. This paper is further extended to introduce colour identification module for NAVI system. Colour information from the interested object is evaluated to determine the colour of the object. The colour output is informed to the blind user through 'verbal voice'.

2. NAVI Hardware

NAVI system consists of tiny vision sensor attached on a sunglass, a stereo earphones and a Single Board Processing System (SBPS) interconnected. SPBS selected for this system is PCM-9550F with Embedded Intel® low power Pentium® MMX 266 MHz processor, 128 MB SDRAM, 2.5" light weight hard disk, two Universal serial bus and RTL 8139 sound device chipset assembled in Micro box PC-300 chassis. The weight of SBPS is 0.7 Kg. Constant 5V and 12 V supply for SBPS is provided from the batteries [6]. Vision sensor selected for this application is a digital video camera, Zion PN615CMOS. Figure 1 shows the processing equipment of NAVI system.

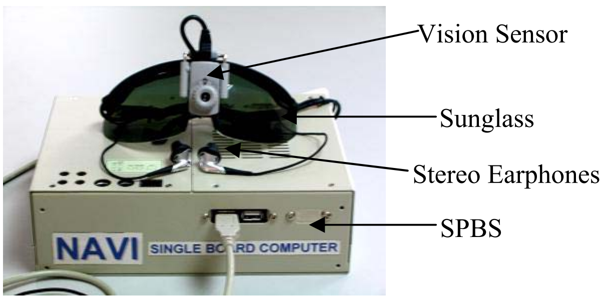

Figure 1. NAVI System

3. Image Processing Methodology

In this paper, improved image processing methodology for NAVI system is presented. The fundamental task of the image processing is to identify and highlight objects in the scene infront of blind user and deliver this information in real time through auditory cues. The term 'real time' is referred to the sampling of the information of the scene infront at the rate of 1 (or 2) image frame per second. In NAVI, duration of sound produced from each sampled image is equal to one second. So, the computational time for image processing has to be less than one second so that the processing of new image and sonification can be carried out during the transmission of previous image's sound [6]. Figure 2 illustrates the block diagram of proposed image processing stages for NAVI. The image processing involves five stages; preprocessing, edge detection, edge linking, object enhancement and noise elimination and object preference procedures. RGB colour image is captured from the vision sensor and used as input image. The usage of colour image provides more information than greyscale image [8]. For preprocessing, the input image is resized to 32 x 32 pixels and contrast stretching technique is applied to each colour component of the resized image. Image resizing is done as to enable faster processing. Contrast stretching is applied to enhance and to correct the uneven illumination in the input image.

From the enhanced image, edges are extracted. The goal of edge detection is to provide structural information about the object boundary. In this work, region inside the closed boundary is considered as object. Thus, by extracting edges in the image, object's feature can be obtained. Several definitions of edge detection on colour image have been proposed [10]. One of the definitions is that, an edge in colour image is exists if edge is present in any one of the tristimulus (i.e. R, G and B) components. In this work, the preprocessed colour image is separated into three components (R, G and B components) and Canny edge detector [2] is applied to each of the colour component. The resulting edges in each component is then combined using logical 'OR' operator.

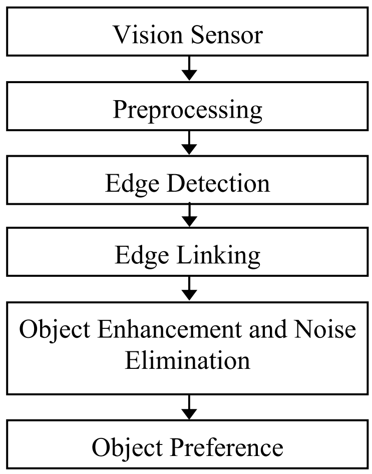

Figure 2. Proposed Image Processing Methodology for NAVI System

In some instances, edges in the image present as a broken segment. Since region inside closed edges are considered as object in this work, it is necessary to close the gaps between these edge fragments. Edge linking is method of assembling edges in the image to form a closed edge. Discontinuity in edge is due to various reasons such as insufficient lighting, geometric distortion and image resizing effect. To connect the edge fragment in the image, edges are scanned in vertical and horizontal direction in a small neighbourhood (3 x 5 pixels for horizontal direction and 5 x 3 pixels for vertical direction). If an edge is present at the specified location and no edges exist between located edge pixels, this non-edge pixel is identified as a candidate edge. The two edge pixels are linked by the candidate edge pixel thus forming an edge link. It is also noted that edge detector fails to detect edge at image border. To connect edges at the image border, labelling is done to the edges to ensure only edges with the same label are connected. First, edges at the row 2 and 31, column 2 and 31 are scanned. A set of rules is formed so that if edges detected in the specified location, pixel on the row 1 and 32, column 2 and 31 will be set to edge pixel. Once necessary edges formed at the border of image, the next step is to connect edges, which have the same labelling. Edge scanning is done to each image border. Starting at the first scanned edge, the end edge is identified if it has the same label with the first edge. All pixels between the scanned edges are converted into edge pixel.

By undertaking the proposed edge linking procedure, closed boundary of objects can be extracted. Then, the object's region pixels have to be enhanced to high intensity and the background's pixels to low intensity. This stage is necessary prior to sonification procedure. During the sonification, the amplitude of sound generated from the image is proportional to the pixel intensity. The image with pixels of high intensity will produce sound of higher amplitude. On the other hand, image with pixels of low intensity will produce sound of lower amplitude. Stereo sound patterns set of pixels with high intensity in the dark background is easy to identify than the low intensity over bright background. It can be felt that the background of the most real world scene is of high intensity than the object. If the image is transferred to sound without any enhancement, it will be a complex task to understand the sound, which is the major problem faced in early works [5]. With this consideration, the object region is enhanced to high intensity. A 'flood-fill' operation is undertaken for the object enhancement [11]. Through experimentation, it is observed that not all edges in the image are object boundaries. These extra edges are considered as noise. Noise has to be eliminated so that exact properties of object can be measured. Morphological operations, erosion and dilation are employed to eliminate noise in the image.

4. Fuzzy Rules Based Object Preference

The main objective of object preference assignment in this work is to identity objects in the image in accordance to human visual preference. In human vision system, a visual consideration is given on an interested object and the other areas are given less consideration. If there is only one object detected in the image, it is considered as object of interest or high-preferred object. But often image in real world scene contain more than one object. By assigning equal preferences to all objects, blind user will have confusion in understanding the features of object. In this work, the central (8 x 8) of image pixel area is called as iris area [4]. In human vision system, the object of interest will be usually at the centre of sight (in this work denoted as iris area). If the object is not within the centre of sight, the object can be brought to the centre by turning the head or by moving the camera. By this assumption, the object of interest has more chances to be in the iris area.

Preference is assigned based on properties such as occurrence in iris area and sizes. Object in the centre is more important for collision free navigation. In industrial vision system, the object of interest is found by known feature of object in the scene. But in NAVI, object is undefined, uncertain and time varying. This is due to the constant shifting of the headgear-mounted orientation by the blind people. To resolve this uncertainty, fuzzy logic is applied. The features to be extracted are described as follows:

a) Size of object - The size of object is the total number of pixels in an object. The size of object, P_1 is given as in equation 1:

$$P_1 = \sum_{i=1}^{32}\sum_{j=1}^{32} O(i,j) \tag{1}$$

b) Object's pixel distribution in Iris Area - the central (8 x 8) of image pixel area is called as iris area. The object's pixel distribution in iris area is expressed as in equation 2:

$$P_2 = \sum_{i=13}^{20}\sum_{j=13}^{20} O(i,j) \tag{2}$$

c) Object's pixel distribution in Outer Area - The object's pixel distribution in outer area is given as:

$$P_3 = \sum_{i=1}^{32}\sum_{j=1}^{32} O(i,j) - \sum_{i=13}^{20}\sum_{j=13}^{20} O(i,j) \tag{3}$$

These features are used as input to fuzzy logic algorithm. Each feature is expressed using three membership functions namely small (*S*), medium (*M*) and big (*B*). Membership functions such as small and big is expressed using trapezoidal curve. For medium membership, a triangular curve is used. The output is the object preference, which has three triangular curve memberships; low (*L*), medium (*M*) and high (*H*). The defuzzification is performed using centroid method. 27 rules are formed based on the human visual consideration stated earlier and some of the rules are given as follows:

- If P_1 is S and P_3 is S and P_2 is *S*, then *O* is *L*
- If P_1 is S and P_3 is S and P_2 is *B*, then *O* is *H*
- If P_1 is M and P_3 is S and P_2 is *M*, then *O* is *M*

The designed fuzzy system produces three outputs, which are low, medium and high preferences. These outputs are referred for assigning the objects with different grey intensity. Dark grey intensity with value of 77 is assigned for low preference object. A medium preference object is shaded to light grey intensity (179) and high preference object to white intensity (255). The processed image I_l is then presented in structural stereo sound pattern for sonification.

5. Sonification

The processed image is transformed into stereo sound patterns [7]. The sine wave with the designed frequency is multiplied with grey scale of each pixel of a column and summed up to produce the sound pattern. The frequency of the sine wave is inversely related to pixel position and the loudness of the sound is made to depend directly on the pixel value of processed image. The sound pattern from each column of image pixels is appended to construct the sound for whole image. The scanning of picture is performed in such a way that stereo sound is produced.

Let

f_o be the fundamental frequency of the sound generator

G be a constant gain.

F_D be the frequency difference between adjacent pixels in vertical direction.

The changes in frequency corresponding to $(i,j)^{th}$ of the pixel in 32 x 32 image matrix is given by

$$f_i = f_o + F_D \quad (4)$$

$$\text{where } F_D = Gf_o(32-i)\,;\ i=1,2,3,\ldots,32 \quad (5)$$

In the proposed system, the frequency is linearly varied, by maintaining F_D as a constant.

The generated sound pattern is hence given by

$$S(j) = \sum_{i=1}^{32} I_1(i,j) \sin \omega(i)t\,;\ j = 1,2,\ldots,32 \quad (6)$$

where,

$S(j)$ is the sound pattern for column j of the image

I_l is the processed image

t = 0 to D; D depends on the total duration of the acoustic information for each column of the image

$\omega(i) = 2\,\pi\, f_i$, f_i is the frequency corresponding to row, i.

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 columns towards the centre.

6. Colour Identification

Colour plays a vital role in human daily life for communication as well as for recognition. Blind people are expected to be interested in the colour of their cloths, the colour of toys and the colour of pictures [13]. The development of colour object identification module for NAVI system is inspired by one prime factor; colour is a powerful descriptor that often simplifies object identification. From the processed image, objects are divided into three preferences. As to avoid confusion to the blind people, the designed colour identification module only aims to identify colour of high preferred object in the image.

Colour can be described as an attribute of visual perception consisting of any combination of chromatic and achromatic content. This attribute can be expressed by chromatic colour names such as red, green, blue etc., or by achromatic colour names such as white, grey, black, etc. Achromatic colour is perceived colour devoid of hue and chromatic colour is perceived colour possessing a hue [12]. Three criteria are identified to be considered for colour module in NAVI system.

- **Functionality –** Human visual system is highly adapted to visual perception of colour. They perform colour recognition simultaneously and effortlessly even in under low illumination. Sighted person can predict more colour and many term can be used to express each colour. In true colour image (24-bit RGB format) a total of 16 million (2^{24}) colours can be generated [3]. Processing all colours is excessive for the blind user. Blind people have to depend on their memory to remember the colour name and to relate each colour to another colour. Therefore, the colour to be identified in NAVI system must be useful and common to blind people.

- **Accuracy –** The colour module should be able to predict colour with minimal error. In general three factors affected the colour perception of an object; surface effect, lighting effect and the sensitivity of vision sensor [3]. Apart from the factors stated, two more factors are considered in this system. It is noted that most of object in real world are in multicolour. The identified colour also affected by the segmented object in the proposed object identification module.

- **Simplicity –** The colour identification module has to be processed in real time. Since processing all colours require a high computational time, colour quantization is needed for NAVI system.

By considering all criteria stated above, in this paper, an algorithm is developed to classify the colours into eleven basic colours. Berlin and Kay [1] state that a total of exactly eleven basic colour categories can be drawn. The eleven basic colour categories are white, black, red, green, yellow, blue, brown, purple, pink, orange and grey. A colour quantization technique is needed to quantitize the input colour into eleven colours. There are many approaches proposed for colour quantization [4, 9]. In this work, an algorithm is designed to classify the colours into eleven basic colours. The colour information of the high-preferred object is scanned from the resized image and the mean colour value for each colour component is obtained using equation 7:

$$\text{colour mean value} = \frac{\sum_{1}^{S_{obj}} C(i,j)}{S_{obj}} \quad (7)$$

where C(i,j) is the colour value and S_{obj} is the area of high preferred object (in pixels).

Achromatic colour such as black, grey and white are determined by comparing the difference in R, G and B value. Usually the differences in these values are relatively small. Chromatic colours are determined by comparing four criteria, which are the maximum mean values, the difference between red and green, the difference between red and blue, and the difference between green and blue. The result from colour detection is transformed into voice output.

7. Experimental Result

Experimentations are done to evaluate the proposed methodology. The set of experiments were conducted to evaluate the image processing procedure using simulated and real-world images. Overall processing time for image processing and colour identification module is less than 1 second. The results of each stage in the proposed procedures for certain indoor and outdoors images are shown in Figure 3. By undertaking the colour edge detection, more information of objects boundary can be detected and extracted accurately. It is observed that objects in the image vary in terms of size, shape and location. The edge linking procedure was able to connect edge fragments in the image. The procedure also connects edges at the image boundary. In this work, object enhancement and noise elimination are undertaken as to impart more consideration to the objects and to eliminate unwanted edges in the image. The morphological operation employs for noise elimination stage also tends to smoothen the object boundary. Designed fuzzy logic system determines the preference/importance of the object from the feature extracted. In general, if the image contains more than one object, only one object in the image is detected as high-preferred object and other objects will be assigned as medium or low preferences. More than one object can have the same medium or low preferences. In experimentation also, a total of 200 objects in real world are tested for colour identification. It is found that the proposed colour identification algorithm was able to identify the colour of object with accuracy of 97 %.

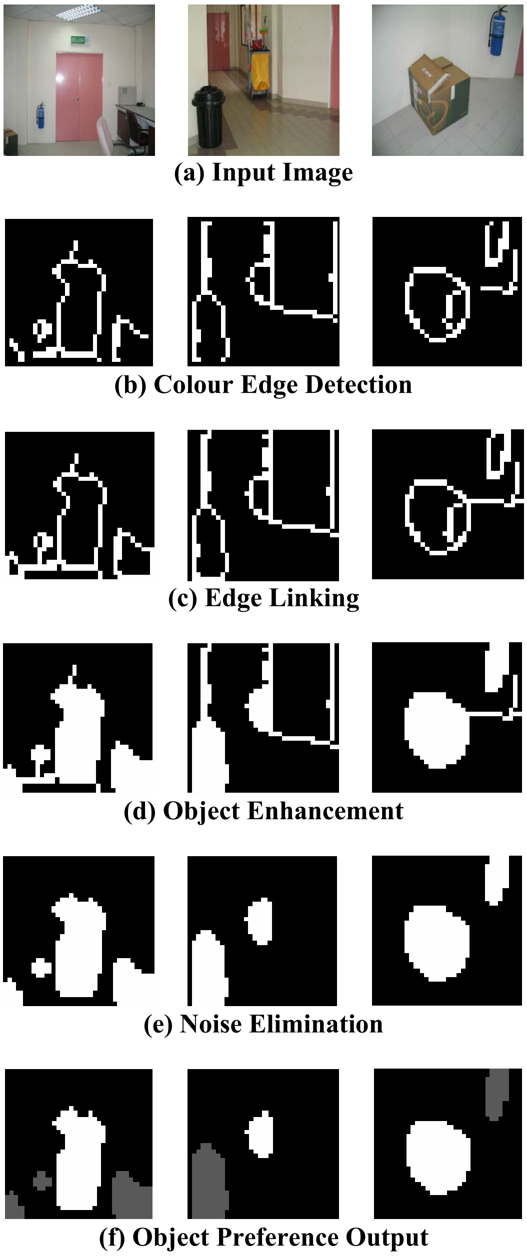

Figure 3. Experimental Results on Image Processing Methodology

8. Conclusions

NAVI is a vision substitute system designed to aid the autonomous navigation of blind people. The system converts image characteristic into stereo sound pattern. This paper presents an improved image processing methodology for NAVI system. In this methodology, colour image is used as input image and image processing procedures is carefully designed so that it can be implemented in real time. The proposed procedure was able to extract objects by identifying the boundary of the objects. Furthermore, noise elimination is undertaken to remove the noise form the image. A visual preferences assignment employed in this work can classify the objects into different intensity. As the processed image is transformed into stereo sound pattern, the sound produced contains less noise and thus more information about the properties of objects can be obtained. In this work, the functionality of NAVI system is also extended by introducing colour identification module. The colour output of the object of interest is presented in verbal voice. With the added functionality, not only the blind user can discriminate the object's properties such as size, shape and orientation but they can also identify the colour of object in their environment. In this work, information regarding the distance of the object from the blind user is not considered directly, it can be developed using stereo cameras [14]. However by comparing the stereo sound patterns from relative distances between the blind person and the object, information regarding the distance of objects can be manipulated by the blind person with experience.

Acknowledgement

Authors wish to thank Universiti Malaysia Sabah for funding this project (B-0401-02-ER/0048).

References

[1] B. Berlin and P. Kay. *Basic Color Terms*. University of California Press, Berkeley, California, 1969.

[2] J. Canny. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, volume 8, pages 679-698, Nov 1986.

[3] P. Green and L. MacDonald. *Colour Engineering: Achieving Device Independent Colour*. Southern Gate, Chichester: John Wiley & Sons Ltd, 2002.

[4] P. Heckbert. Colour Image Quantization for Frame Buffer Display. In *Proc. of the 9th Annual Conference on Computer Graphic and Interactive Techniques*, pages 97-307, 1982.

[5] P.B.L. Meijer. An Experimental System for Auditory Image Representations. *IEEE Transactions on Biomedical Engineering*, volume 39, No. 2, pages 112-121, Feb 1992.

[6] R. Nagarajan, S. Yaacob and G. Sainarayanan. Fuzzy Based Human Vision Properties in Stereo Sonification System for Visually Impaired, In *Proc. of SPIE – Intelligent Systems and Advanced Manufacturing*, The International Society for Optical Engineering, volume 4572, pages 525-534, Nov 2001.

[7] R. Nagarajan, G. Sainarayanan and S. Yaacob. Role of Object Identification in Sonification System for Visually Impaired. In *IEEE TENCON-International Conference*, pages 735-749, Bangalore, India, October 2003.

[8] C. L. Novak, and S. A. Shafer, Color Edge Detection. In *Proc. DARPA Image Understanding Worksop*, volume 1, pages 35-37, 1987.

[9] S.C. Pei and Y.S. Lo. Color image compression and limited display using self-organization Kohonen map. *IEEE Transactions on Circuits and Systems for Video Technology*, volume 8, issue: 2, pages 191 -205, April 1998.

[10] W.K. Pratt. *Digital Image Processing*, 3rd ed. New York, John Wiley & Sons, 2001.

[11] J. C. Russ. *The Image Processing Handbook*, 2nd ed. Boca Raton, Florida: CRC Press, 1995.

[12] N. Sugano. A Fuzzy Set Theoretical Approach for Color Specification Using Color Names. In *Proc. of IEEE International Conference on System, Man, and Cybernetics*, pages 230-235, 1999

[13] R. Walhof. Color. The Braille Monitor, volume 43, No. 12. December 2000. http://nfb.org/BM/BM00/BM0012/bm001210.htm

[14] F. Wong, R. Nagarajan and S. Yacoob. Application of Stereovision in a Navigation Aid for Blind People. In Proc. of fourth International Conference of Information, Communications and Signal Processing and Fourth IEEE Pacific-Rim Conference on Multimedia, Paper No. 2A2.3 (CD-ROM), Singapore, 2003.

[15] F. Wong, R. Nagarajan, S. Yacoob, C. Ali and B. Nour Eddine. Electronic Travel Aids for Visually Impaired – A Guided Tour. In *Proc. of Conference in Engineering in Sarawak*, pages 377-382, 2000.

[16] World Health Organization (WHO), World Sight Day: 10 October 2002.