International Journal of Computational Intelligence Techniques
ISSN: 0976-0466 & E-ISSN: 0976-0474, Volume 3, Issue 2, 2012, pp.-102-106.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000221

ACOUSTIC VISION - ACOUSTIC PERCEPTION BASED ON REAL TIME VIDEO ACQUISITION FOR NAVIGATION ASSISTANCE

RAO S.K.[1]*, PRASAD A.B.[1], SHETTY A.R.[1], CHINMAI[1], HEGDE R.[1] AND BHAKTHAVATHSALAM R.[2]
[1]Department of Telecommunication Engineering, BMS College of Engineering, Bangalore-560 019, Karnataka, India.
[2]Super Computer Education and Research Center (SERC), Indian Institute of Science, Bangalore-560 012, Karnataka, India.
*Corresponding Author: Email- supreethkrao@gmail.com

Received: October 25, 2012; Accepted: November 06, 2012

Abstract- A smart navigation system based on an object detection mechanism through real time video processing has been designed to detect the presence of obstacles that impede its movement. This paper is discussed keeping in mind the navigation of the visually impaired. A PAL video camera feeds image frames pertaining within its vicinity at a rate of 30 fps to a *Da-Vinci Digital Media Processor*, DM642. The processor carries out image processing techniques on every frame and the result contains information about the object in terms of image pixels. The algorithm aims to select that object, among others detected in the frame, which poses maximum threat to the navigation. The selected object then translates to one out of three sounds, whose pitch informs navigator about the obstacles' relative threat. This paper implements a more efficient algorithm compared to its predecessors.

Keywords- DM642, NAVI.

Citation: Rao S.K. et al. (2012) Acoustic Vision - Acoustic Perception Based On Real Time Video Acquisition For Navigation Assistance. International Journal of Computational Intelligence Techniques, ISSN: 0976-0466 & E-ISSN: 0976-0474, Volume 3, Issue 2, pp.-102-106.

Introduction

The visual information is the basis for almost all the navigational tasks and hence visually impaired people are at disadvantage as the appropriate information about the environment is not available to them [1]. An Electronic Travel Aid (ETA) is a form of assistive technology, which enhances the mobility for blind [2]. Assistance for the blind or visually impaired can range from simple measures, such as a white cane or a guide dog, to a very sophisticated computer technology (enhanced imaging, synthetic speech, optical character recognition, etc.). Many of those who are visually impaired can maintain their current employment or be trained for new work with the help of such aids. This paper deals with a vision substitution system that is based on an image to sound conversion concept. This finds particular applications for the navigation of the visually impaired and even in the case of self-localizing autonomous rovers. Output of this system can either be used to actuate a smart control system, or transformed into an audio signal which a blind user can interpret and thus navigate. The paper aims at creating a portable system that allows visually impaired individuals to travel through familiar and unfamiliar environments without the assistance of guides. This paper is organized as follows. Section II deals with the related work carried out. Section III presents the proposed algorithm while Section IV explains the methodology. Section V discusses the hardware implementation. Section VI presents the results and their interpretations. This paper is then concluded in section VII.

Related Work

Nagarajan R., et al [3], have developed "Navigation assistance for visually impaired (NAVI)" which is a vision substitute system designed to assist blind people for autonomous navigation. Fernandes H., et al [4] proposed a paper that focused mainly in the development of a computer vision module for a Smart-Vision system. G Balakrishnan et al [5] proposed a system for visually impaired which utilizes stereo vision, image processing techniques and a sonification procedure to support navigation for blind. Sandra Mau et al developed an ETA for blind people which consists of an RFID reader carried by the user and a network of low cost RFID tags in the building to be navigated [6]. Xu Jie et el proposed an ETA system which helps the blind people to judge current environment. The proposed method uses image processing techniques to get the information of the blind sidewalk orientation [7]. Faria, J. et al [8] proposed the development of an electronic white cane which helps the visually impaired persons to navigate in both indoor and outdoor environments which provides geographical information using RFID technology. João José et al [9] developed a Smart Vision system for detecting path borders and the vanishing point which enables blind persons to correct their heading direction on paths. A low cost system for blind persons was proposed by D Bernabei et al [10], which takes as input, the depth maps produced by the Kinect © device coupled with the data to provide a registered point based 3D representation of the scene in front of the user. A 3D navigation system for the blind using a pair of glasses

equipped with cameras and sensors was developed by Parisian researchers [11].

Proposed Algorithm

A vision acquisition device, in this case a Phase Alternating Line (PAL) video camera is used as the "eye" of system. Image frames are procured from the video and subjected to a series of image processing techniques. Block diagram giving an overview of the algorithm used for the system implementation is presented in [Fig-1]. Location and size are the two object parameters that are given weightage in the algorithm, for threat detection. A flood function has been designed to calculate these parameters. Information from the flood function is then used to categories the objects by assigning priority values, based on their proximity and size. Object that has gained the highest priority is selected. Based on pre-defined thresholds, acoustic transformation is performed resulting in one out of three sound signals (No sound being one of them). The pitch of sounded signal gives the ultimately intended idea about the surroundings, which helps the user to have a collision free navigation. Object recognition is being implemented for the next version, to inform the obstacles' identity and thus better present the surroundings to the blind user.

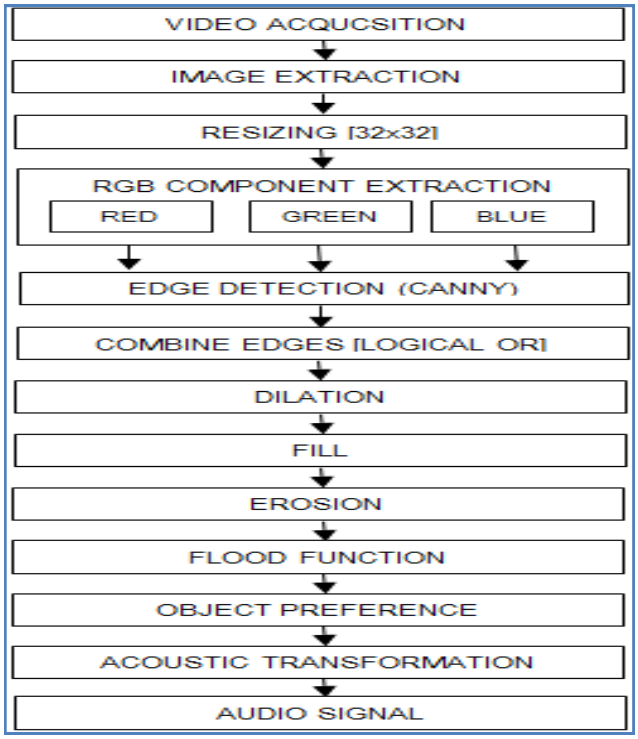

Fig. 1- Proposed Algorithm

Methodology

Proposed methodology involves the following pre-processing techniques

Resizing

The captured image is first resized to 32x32. This achieves the real time processing pre-requisite of small computational time, while it also provides flexibility in choosing the camera fairly independent of its resolution. The above size is selected keeping the data loss resulting from resizing, into consideration.

Edge Detection

The objects in the image are detected by their boundaries through Edge detection, which is a technique that extracts edges transitions in an image. Amongst the various types of edge detection- Sobel, Canny, Prewitt, Laplacian to name a few, our algorithm responds well to both Canny and Sobel. Image acquired by the camera being a color image in our case, contains red, green and blue (RGB) components [12]. The use Color image segmentation results in greater accuracy as against using Grayscale edge detection [13].

Dilation

Morphological processing involves operations that process images based on shapes. They apply a structuring element to the input image that suitably altering it. Dilation and erosion are the two morphological operations that the algorithm uses. Hence, we extract the RGB color components of the resized image and perform edge detection on each of the components. A single final edge detected binary image is obtained by performing the logical OR operation on the three edge detected images. This image consists of clearly defined objects whose boundaries are in white.

Fill

As it is required to calculate the size of the object, the area within each object must be obtained i.e., the number of pixels that make up the object. The image as of now consists of objects outlined in white. The fill function, when applied to the above image, fills in the black area within objects with white pixels. It can thus be inferred that the object size is the number of white pixels it contains.

Erosion

While dilation adds pixels to the boundaries of objects in an image, erosion removes pixels based on the structural element defined. This sharpens the dilated objects in addition to removing unwanted white specks, which could otherwise be interpreted as small objects themselves. The structural element chosen here is a disk with radius being one.

Proximal Area

The location of the objects is an important aspect for collision free navigation. The presence of objects in certain parts of the image poses greater danger to the navigator than their presence in other parts. To understand this, let us divide an image into four parts, left, right, center and back, as shown in [Fig-2].

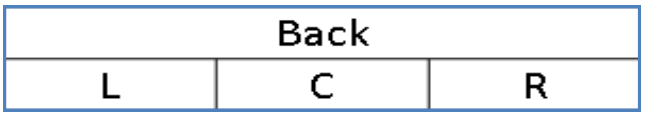

Fig. 2- Analysis of the image frame

Consider the following cases in [Fig-3] with the assumption that the user is walking straight ahead:

Case (a): An object is in the left (L) portion of the image and it is apparent that this object does not cause any obstruction to the navigator.

Case (b): The object in the right (R) portion of the image. This too, will not obstruct the path.

Case (c): Here, the object is at the back portion of the image. As we can see, the user can walk a short distance before he encounters this object.

Case (d): In this figure, however, since the object is present at the central portion of the image, we can see that it poses an immediate obstruction to the navigator. Therefore, we can infer that this region of the image should be given higher preference when compared to the other regions.

Fig. 3- Selection of Proximal Area

To classify the objects based on their location, the frame has been divided into three regions [Fig-4], two of which constitute the high priority region-Proximal area consisting of A1 & A2 [Fig-2]. Within the Proximal area, objects present in A1 are assigned the highest priority values, while those present in A2 are assigned a lower value of priority.

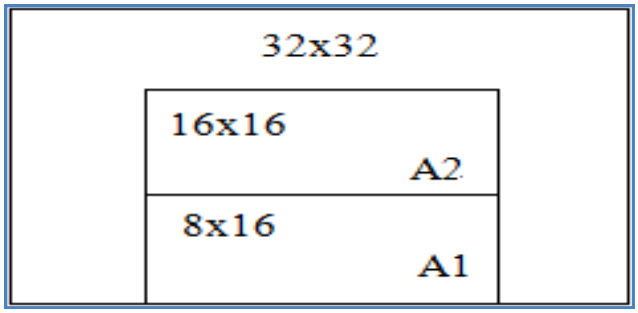

Fig. 4- Proximal Area

In this paper, we assign preferences to A1 and A2 through the use of masks M1 and M2 consisting of ones. M1 is a mask of dimension 24x16 highlighting the objects in the areas A1 and A2. M2 is a mask of dimension 8x16 highlighting the objects in A1only. Using these masks, the following operations are performed:

M1 = Image & M1; (objects in A1 and A2 are highlighted)

M2 = Image & M2; (objects in A1 alone are highlighted)

Image = Image + M1 + M2;

This results in an image consisting of 1s, 2s and 3s in the regions outside the proximal area, in A2 and in A1 respectively.

Flood Function

The objects in the image consist of 1s, 2s and 3s; hence, calculating the object size would be to calculate the total number of 1s, 2s and 3s. Also, the concentration of objects (which is the number of pixels) in A1 and A2 have to be calculated. This would mean counting the number of 2s to calculate object concentration in A2, and 3s to calculate the object concentration in A1.

In order to calculate the size of the objects present in the frame, and their concentration in A1 and A2, a flood function has been designed.

The image is first searched for a one, two or three, and the flood function is called when these pixel intensities are encountered. The function called at a pixel spreads to its neighbors if its pixel intensities are not zeros. The function has been defined for eight connectivity, i.e., all eight neighbors of each pixel are checked for 1, 2 or a 3. In effect, when a flood function is called at a pixel of an object, then it spreads to cover the entire object till it reaches the object's boundaries. During flooding, pixel count can be incremented (which would be the object size) and also, the number of twos and threes can be counted (which would mean the object concentration in A2 and A1 respectively). To prevent the function from flooding into already explored pixels, the intensities of those pixels where the function has been called are changed to zero. Therefore, not only does the function count required values, but it also shrinks the object into inexistence. This has no consequences as all necessary information has already been gathered.

The scanning continues and the procedure is repeated as and when other objects are encountered. A database of object sizes and concentrations is thus created.

Priority Assignment

Objects falling in the Proximal Area should be given high priority. Not only should this been done, but importance should also be given to the object size. Consider the following conflicting cases:

Case 1: Suppose there are two objects lying in A2, the larger object should be given a higher priority.

Case 2: Suppose there is a large object in A2 and a smaller object in A1, as the user encounters the object in A1 first, objects in A1 should be given more priority than those in A2, regardless of the difference in size.

Case 3: Suppose a small percentage of a huge object lies within A1and A2 and an object of smaller size lies only within A2, then that object whose concentration within A1 or A2 is more gains higher priority.

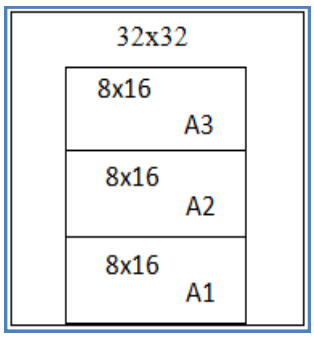

Fig. 5- proposed proximal area

Although the proposed proximal area in [Fig-4] serves our purpose for practical considerations, in this section, we are proposing an

alternate proximal area geometry as shown in [Fig-5], which when used in the algorithm- will improve the accuracy of the system to a finer resolution. The algorithm will follow the same approach in interpreting the objects detected; only difference being that it assigns higher priority to objects in A2 than in A3 now. Hence, the priority assignment will be in the following order- p(A1)>p(A2)>p(A3). This would benefit the system's accuracy in the following case (please refer [Fig-4]) - There is no object in A1; Two objects- obj1 (larger and present farther) and obj2 (smaller yet a significant threat, and is present nearer to the user) are present in A2. This would ideally result in the detection of just the larger of the two objects as there is an ambiguity in its position (that is both are present in A2) according to the algorithm. However, as mentioned before, the proposed image frame segregation in [Fig-4] gives fairly accurate results in practice and hence we have chosen to use this geometry for this paper.

Acoustic Transformation

Each object by now would have gained a priority value. Among all other objects detected, that object which has gained the highest priority must have its presence conveyed to the blind user. Thus, the algorithm translates the object into an audio signal- a beep. All priority levels are categorized into three sounds:

i. A sound of high pitch indicating the presence of an object posing greatest threat and implying that the user must immediately change his direction

ii. A sound of medium pitch indicating that an object would soon pose the greatest threat if the user continues on his path and

iii. No sound, which informs that there is no object yet which can pose a threat.

Hardware implementation

The system has used the Digital Media Processor TMS320DM642 (Version 3), which belongs to the Da Vinci family of Texas Instruments' C6000 series. The DM642 Evaluation Module (EVM) is a low-cost, high performing video & imaging development platform designed to jump-start application development and evaluation of multichannel, multi format digital and other future proof applications [14]. DM642 has been specially designed for real time video and audio processing, with dedicated video encoders and a decoder [15]. Leveraging the high performance of the TMS320C64x DSP core, this platform supports TI's TMS320DM642, DM641 & DM640 digital media processors [16]. The TMS320C64x™ DSPs (including the TMS320DM642 device) are the highest-performance fixed-point DSP generation in the TMS320C6000™ DSP platform. TMS320DM642 device is based on the second-generation high-performance, advanced VelociTI™ Very-Long-Instruction-World (VLIW) architecture (VelociTI.2™) developed by Texas Instruments (TI), making these DSPs an excellent choice for digital media applications. TMS320DM642 offers a speed of 720MHz, 4 MB Flash, 32 MB of 133 MHz SDRAM and 256 kbit I2C EEPROM [15]. The JTAG emulator used to communicate with the processer is XDS510USB Plus. A PAL (Phase Alternating Line) camera has been utilized to acquire the input (at 30 frames per second) to the system and a set of 3.5 mm jack ear-phones have been used to provide the output of the system to the user. [Fig-6] shows the hardware implementation of the system.

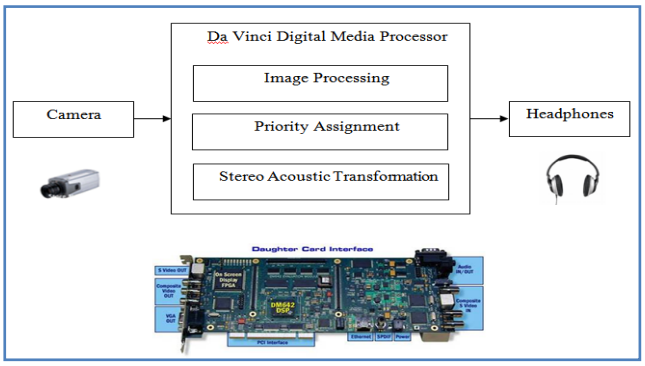

Fig. 6- Hardware Implementation using DM642

Results and Discussion

This section shows the results for a sample image as seen in [Fig-7] along with observations following [Fig-8]. The results depict the presence of two objects in the image, the object 1 is of size 114 and object 2 is of size 9. 9 pixels of object 1 is present in the A1 region of the frame and the remaining 105 pixels are present in the A2 region of the frame whereas object 1 does not lie in either A1 or A2. Thus the highest priority is given to object 1 and the audio output is sounded accordingly.

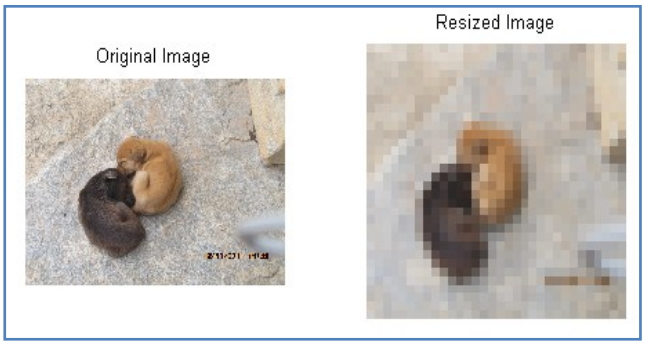

Fig. 7- Image Extraction & Resizing

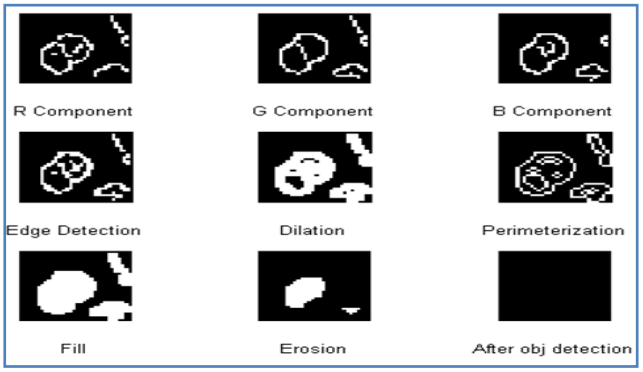

Fig. 8- Image processing steps

Table 1- End results considering five objects

A1	9	0	0	0	0
A2	105	0	0	0	0
Objsize	114	9	0	0	0
Priority	115	0	0	0	0
Highest Priority	115				

[Table-1] shows the end results considering a maximum of five objects.

Conclusion

Acoustic vision is a sensory substitution system (vision) that acquires, processes, analyses and understands images from the real world and ultimately provides a synthetic vision through sound, that is relevant to obstacle detection based navigation. Here, image extraction from a video input is carried out; boundaries of objects present in the image are identified and the objects' sizes are calculated. Importance is given to the size and proximity of objects through well-defined iris areas. The flood function counts the size of objects and their concentration in these iris areas. Assignment of priorities then takes place to categorize detected objects based on the amount of threat they are likely to pose; finally, an acoustic transformation is performed to translate the visual information to an interpretable audio beep. This is executed by the Da Vinci Digital Media Processor, DM642, in half a second, and the real time system's execution dives again to the first step and the procedure is carried out repeatedly.

Neuroscience and psychology research indicate recruitment of relevant brain areas in seeing with the use of sound, as well as functional improvement through training. However, the extent to which the cortical plasticity allows for functionally relevant rewiring or remapping of the human brain still remains majorly unknown while it is being investigated in an open collaboration with research partners all around the world.

In effect, this system translates one sense organ's experience to be understood by another which eventually gets used to the user, owing to the neural plasticity of the human brain. The visually impaired persons can now "see" by listening to the output of this algorithm.

Acknowlegement

We are ever indebted to our college, BMS College of Engineering, for providing us with a healthy environment conducive to learning that enabled us to conceptualize and conduct a work such as this.

We would like to thank Dr. B. Kanmani, Head of the department, Telecommunications Engineering, BMSCE, Bengaluru, under whose headship and encouragement, we were given access to a very well-equipped lab, without which this work would have just remained in a book of ideas.

We would like to express our sincere gratitude to Mr. Gowranga, SERC, Indian Institute of Science, Bengaluru, for his timely inputs that helped us progress towards our goal in the right direction.

Our deepest thanks to Mr. Jayaramudu, Cranes International Software Ltd., Bangalore, for his invaluable help in setting up the DSP DM642 Processor.

References

[1] Santhosh S.S., Sasiprabha T., Jeberson R. (2010) *International Conference on Recent Advances in Space Technology Services and Climate Change*, 277-282.

[2] Mounir Bousbia-Salah; Mohamed Fezari Rachid Hamdi (2005) 16*th IFAC World Congress,* 1401.

[3] Nagarajan R., Sainarayanan G., Yacoob S., Porle R.R. (2004) *IEEE Region 10 Conference*, 455-458.

[4] Fernandes H., Costa P., Filipe V., Hadjileontiadis L., Barroso J. (2010) *World Automation Congress,* 1-6.

[5] Balakrishnan G., Sainarayanan G., Nagarajan R. and Sazali Yaacob (2006) *International Journal of Information and Communication Engineering*.

[6] Sandra Mau, Nik A. Melchior, Maxim Makatchev, Aaron Steinfeld (2008) *Technical Report*, Pittsburgh, Carnegie Mellon University.

[7] Bin Ding, Haitao Yuan, Xiaoning Zang, Li Jiang (2007) *International Conference on Wireless Communication, Networking and Mobile Computing*, 2058-2061.

[8] Xu Jie, Wang Xiaochi, Fang Zhigang (2010) *International Forum on International Technology and Applications*, 2, 431-434.

[9] Faria J., Lopes S., Fernandes H., Martins P., Barroso J. (2010) *World Automation Congress*, 1-7.

[10] Ran L., Helal S., Moore S. (2004) *Second IEEE Annual Conference on Pervasive Computing and Communications*, 23-30.

[11] João José, Miguel Farrajota, João M.F. Rodrigues, J.M. Hans du Buf (2011) *International Journal of Digital Content Technology and its Applications,* 5(5).

[12] Hui Tang, Beebe D.J. (2006) *IEEE Transactions Neural Systems and Rehabilitation Engineering*, 14(1), 116-123.

[13] Velazquez, Pissaloux E.E, Guinot J.C. (2005) *27th Annual International Conference of the Engineering in Medicine and Biology Society,* 6821-6824.

[14] Bernabei D., Ganovelli F., Di Benedetto M., Dellepiane M. and Scopigno R. (2011) *International Conference on Indoor Positioning and Indoor Navigation*, Guimara˜ Es, Portugal.

[15] Paul Bach-y-Rita and Stephen W. Kercel (2003) *Trends in Cognitive Sciences,* 7(12), 541-547.

[16] Soumya Dutta, Bidyut B. Chaudhuri (2009) *International Conference on Advances in Recent Technologies in communication and Computing,* 337-340.

[17] Sapna Varshney S., Navin Rajpa and Ravindar Purwar (2009) *International Conference on Methods and Modelsin Computer Science*.

[18] Huska J., Kulla P. (2011) *Advances in Electrical and Electronics Engineering,* 31-34.