A PDA Based Artificial Human Vision Simulator

Jason Dowling
School of Engineering Systems.
Queensland University of Technology, Brisbane, Australia
email j.dowling@qut.edu.au

Anthony Maeder
E-Health Research Centre.
CSIRO, Brisbane, Australia
email Anthony.maeder@csiro.au

Wageeh Boles
School of Engineering Systems.
Queensland University of Technology, Brisbane, Australia
email w.boles@qut.edu.au

Abstract

Much recent research attention has focused on providing some form of visually meaningful information to blind people through electrical stimulation of a component of the visual system. Current technology limits the number of perceived points of light (phosphenes) that can be provided to a user and methods are required to optimize the amount of presented information. This paper describes a PDA based artificial human vision simulator, and proposes a method for alerting a user of possible looming obstacles. Experimental results indicate that obstacle alerts can be successfully provided, however with the current simulator components, high-quality lighting and accurate image segmentation is critical for reducing the number of false alerts.

Keywords

Visual prostheses, blind mobility, artificial human vision, image processing, simulation.

INTRODUCTION

Existing mobility aids for the blind typically provide mobility information via tactile (eg. long cane or guide dog) or auditory (eg. ultrasound based aids) sensation. An alternate approach is to provide a vision substitute by electrically stimulating a component of the visual system. This approach is referred to as Artificial Human Vision (AHV) or a “visual prosthesis”. During electrical stimulation a blind person may perceive spots of light, called “phosphenes”. Currently four locations for electrical stimulation are being investigated: behind the retina (subretinal), in front of the retina (epiretinal), the optic nerve and the visual cortex (using intra and surface electrodes) [1]. As there are technical limits to the number of electrodes which can be implanted, image processing techniques are required which can maximize the usefulness of the available phosphenes.

As blind mobility aids are often expensive and require extensive training, it is desirable to be able to objectively compare the usefulness of different devices. Psychophysical and mobility course assessment should help in developing and comparing AHV systems with other technical aids for the blind. Due to the difficulty in obtaining experimental participants with an implanted AHV device, a number of simulation studies have been conducted with normally sighted subjects. The simulation approach assumes that normally sighted people are receiving the same experience as a blind recipient of an AHV system.

The first reported AHV simulation research was conducted by Cha et al. [2] at the University of Utah, who built a device consisting of a video camera connected to a monitor in front of the subject’s eyes. A perforated mask was placed on the monitor to replicate the effect of individual phosphenes. This research found that a 25x25 array of simulated phosphenes, with a field of view of 30^{o} would be required for a successful device.

The simulation display in Cha et al. (1992) used a simple television-like display. A more sophisticated approach was proposed by Hayes et al. [3]. In their research, two different image processing applications were used to display simulated phosphenes to a seated subject, who wore a head mounted display. Phosphenes were presented as solid grey scale values equal to the mean luminance of the contributing image pixels or as a dome-shaped gray-scale distribution whose centre had the mean luminance of the contributing image pixels, and the edges matched the background intensity, The main result was to conclude that the phosphene array size will be the most important factor in a useable prosthesis.

Another image processing approach has investigated the requirements for AHV facial recognition [4]. Consisting of a Low Vision Enhancement System (LVES) connected to a PC, the simulation displayed a circular ‘dot mask’ to match an ideal prosthesis output. Electrode properties (such as drop outs; size and gaps), contrast and gray levels could be varied experimentally. The authors reported that reliable face recognition using a crude pixelized grid can be learned and may be possible even with a crude prosthesis.

Static simulation image research has also been conducted by Boyle et al. [5], who found that most image processing techniques were not very helpful at low resolutions (typically a 25x25 array).

With the exception of the research by Cha et al. [2] , the simulation studies described have involved static images. However the ecological approach to perception, widely referenced in the literature on blind mobility, emphasizes movement in a complex and changing environment [6]. Our current research at QUT is investigating methods for enhancement of mobility for AHV system users using image sequences. This research has suggested that the display from a visual prosthesis could use different information reduction and scene understanding information methods depending on the task context and the type of scene. For mobility purposes this display depends on three main dimensions of the current scene (Figure 1): The *Context* (we may need more information reduction in a cluttered shopping mall than street crossing); the *Task* (safely negotiating a traffic crossing may require different information than finding a doorway) and active *Alerts* (the system should provide a warning in hazardous situations) [7].

A visual prosthesis simulation has been developed to investigate the mobility display framework shown in Figure 1, This portable head mounted device consists of a Personal Digital Assistant (PDA) and an attached digital camera. The PDA display is used to present the phosphene simulation. A normally sighted subject can wear the device and be assessed on various mobility tasks under different contexts, alert scenarios and image processing conditions. A sheet of material (not shown in Figure 1) is used to limit the subject's visual information to the PDA display.

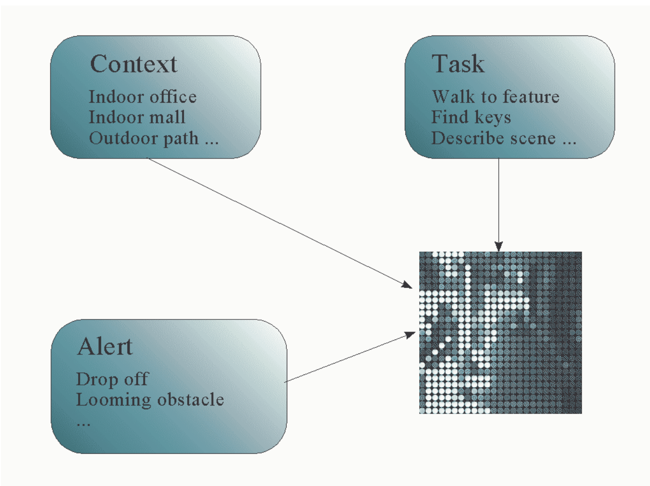


Figure 1. Proposed mobility display framework

HARDWARE

The main benefit of using a PDA is the small size, lightweight and a lack of connecting cables. Current generation PDA's are however constrained by relatively slow CPU and bus speeds, and lack a floating-point unit for real number computation.

The current project uses a HP iPaq 2210 Pocket PC that includes an Intel XScale PXA255 (400 MHz) processor and has an internal bus speed of 200MHz. For image capture, a Lifeview Flycam CompactFlash Camera Card is used, consisting of a 350K CMOS sensor, with a viewing angle of 52°. The combined weight of the camera and PDA is 164 grams.

We have adapted a standard headgear device to include a bracket for holding the Pocket PC in front of a subject's eyes (Figure 2). The viewing distance from subject eyes is approximately 65 cm. The PDA screen display is 8.89cm diagonal with a resolution of 240x320 pixel.

Figure 2. Front and side views of the AHV simulator used in the present study.

SOFTWARE

The main requirement for the simulation software is to convert input from the camera into an on-screen phosphene display. The current system reduces the resolution of captured images from 160x120 RGB to 32x16 or 16x12 greyscale "phosphenes". In addition, background processing need to determine if an alert warning should be displayed.

The Flycam-CF Software Development Kit was used for accessing images from the camera. The simulator software was developed in Microsoft embedded Visual C++ version 4.0. A 32 bit Windows test application was also developed using Microsoft Visual C++ version 6.0 to test methods on image sequences previously captured from the PocketPC and camera.

The traditional approach to image based obstacle avoidance, using a single camera, is to estimate the optical flow within the image sequence, compensate for camera motion (ego motion), and suggest turning towards the direction where the optical flow is smaller [8]. However the calculation of optic flow and ego motion is computationally expensive, particularly on a PDA. The approach used in the current project is to segment each image, and then check the size and rate of expansion of each segment between contiguous images. To improve computation time, each 5x5 pixel area from the original 160x120 pixel image is used to generate one 32x24 phosphene "blocks".

The main steps used in the PDA simulation are shown in Figure 3. A set of arrays for both the current and previous image is maintained, including the block grey-level value, warning segments, and segment size. An array of allocated segments is also maintained across images.

Steps 1-4

Initially each 160x120 pixel RGB bitmap supplied by the camera SDK is converted into a 256 grey-level image. If the difference between the sum of grey-level values in the current image and the sum of grey levels in the previous image is greater than a threshold, the current scene has assumed to have changed and the previous and segment arrays are reset (step two). The threshold used is 245760, chosen as a 10% change in total image grey level for the image: (160x120x128)/10).

In step three, the 256 level image is converted to an 8 grey-level array. This reduction of grey-level information assists with the execution speed of image segmentation. The 3x3 median filter, applied in step four, is applied to reduce noise. This filter is computationally efficient, as there are only 8 grey levels to consider.

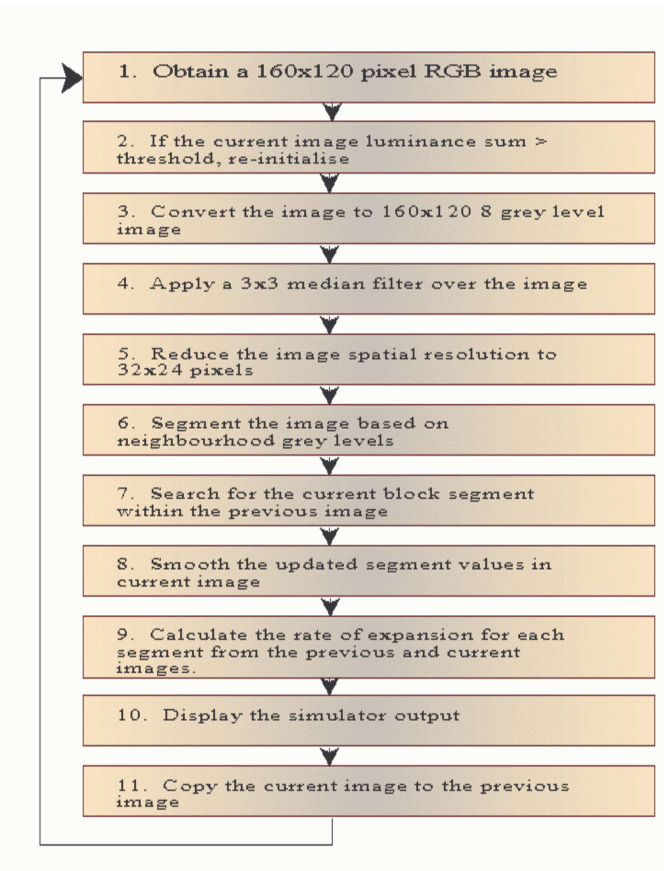


Figure 3. Block based AHV simulation steps.

Step 5

In this step the 32x24 "block" array is generated. The value of each "block" is determined by calculating the median value of the 25 contributing pixels in the original 160x120 image. Image segments that are expanding at a certain rate and larger than a certain size are used to determine the presence of a looming obstacle: therefore the loss of spatial resolution is compensated by improved search time in the segmentation steps.

Step 6

Steps 6 through 10 use the 32x24 block array. The eight neighboring blocks of each block are scanned in a clockwise manner for a matching grey-level value. If any of the grey-level values match, and the matching block has been allocated to a segment, the current block segment is set to the matching block segment. If there is no matching grey-level or segment available, a new segment is allocated.

Step 7

This step searches for the position of each current block array element in the block array created from the previous image. As the camera is moving between frames (due to head movements and gait), ego motion is considered by searching over a 5x5 block area in the previous block array in the following manner: the current block value is first compared against the previous block array value. If there is no match, a search is conducted over the neighboring 8 blocks in the previous block array. If a match is still not made, a search is conducted over the 16 blocks neighboring the 8 blocks. If there is no match from any of the 25 blocks, a new segment is allocated to the current block.

Step 8

The final stage of segmentation stage smoothes the current block array segments. For each block, a search is performed on the immediate 8-block neighborhood and, if there is a matching grey-level value, the current segment is updated to the matching block's segment.

Step 9

To check the rate of expansion, a comparison is made between the area (number of blocks) of each segment. Segments that are larger than a preset threshold (currently 20 blocks in area) are considered. If the rate of expansion (Current image allocated segment size/Previous image allocated segment size) is greater than a threshold, an alert is set for that segment.

Steps 10-11

Finally the "phosphenes" are displayed on the PDA display. If a segment has been identified as an alert, the segment blocks are identified with an "alert colour" (currently pink). As the Pocket PC operating system does not support the Microsoft DirectX set of APIs for high performance graphic display, the Game Application-Programming Interface (GAPI) is used to directly access the display memory. In our simulation display, the block array is expanded to fill the 240x320 pixel display. To improve efficiency, blocks are only displayed if they differ from the previous display.

Sample processing

Figure 4 illustrates the algorithm steps on a single image. In this image sequence a subject has veered into bushes next to a path. 4a is the original 160x120 pixel grey scale image. 4b is the same image after median filtering and conversion to 8 grey level values. 4c is the 32x24 block representation of 4a. 4d shows the location of alert segments which have been set for this image.

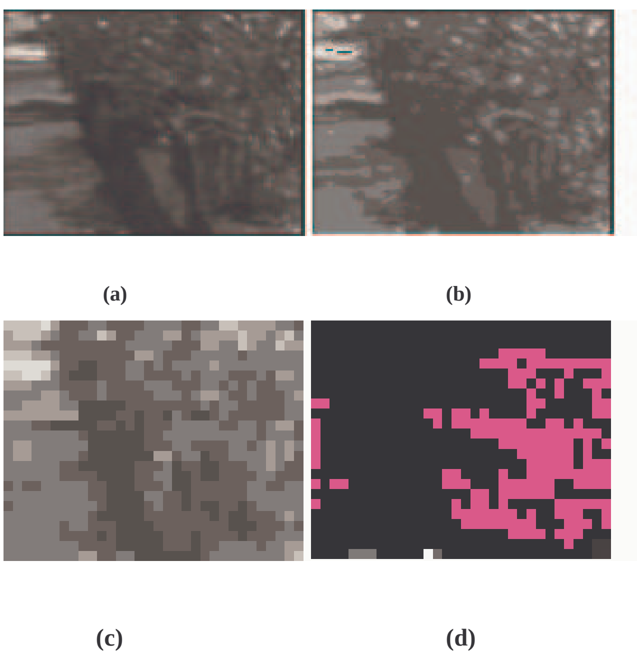


Figure 4. Example of block based image processing

EXPERIMENTAL RESULTS

To evaluate the performance of the obstacle alert component of the AHV simulation, two sets of image sequences were captured at different times of the day using the simulation held at head height. The first sequence involved walking slowly around a bend and towards a postal box (approximately 15 metres in total). In the second sequence, the experimenter walked towards a bus shelter obstacle along a path with overhanging trees (a distance of approximately 10 metres). Both sequences ended with the collision of the camera and the final obstacle. These image sequences were then analyzed on the PC based version of the alert software. Alerts were compared against identified obstacles within the sequence (fences, overhanging trees, etc).

The results for the postal box (Table 1) sequence are influenced by a white fence on one side of the path. During the sequence captured at early afternoon, this fence was captured less frequently which led to a reduction in valid alerts. This suggests that following known structures, such as walls or fences, may be a useful method of using an AHV system (a similar method, called shorelining, is frequently used by blind people while walking next to walls or paths). Aside from the early afternoon sequence, the ratio of correct/total number of alerts (Figure 5) decreased as the experimenter moved away from the fence and increased again towards the postal box. An example of correct obstacle identification for the mid-morning postal box sequence is shown in Figure 8.

Table 1. Postal Box image sequence results.

Time of Capture	Mean Grey level	Correct Alerts	Total Alerts	Result (%)
Mid morning	110.60	13	18	72.2
Early afternoon	102.05	7	19	36.8
Mid afternoon	91.91	14	21	66.6
Late afternoon	87.70	11	23	47.8

Table 2. Bus shelter image sequence results.

Time of Capture	Mean Grey level	Correct Alerts	Total Alerts	Result (%)
Mid morning	72.73	7	8	87.5
Early afternoon	110.48	18	18	100.0
Mid afternoon	76.44	3	7	30.0
Late afternoon	81.82	1	4	25.0

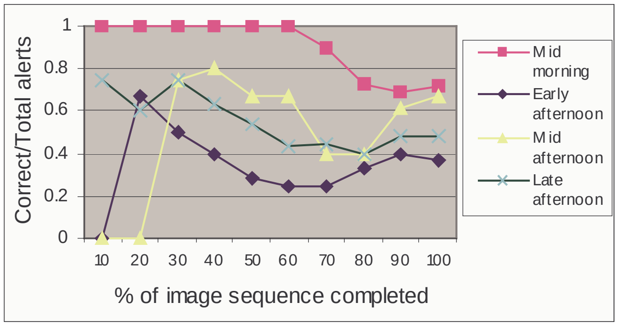


Figure 5. Alert ratios for each postal box sequence.

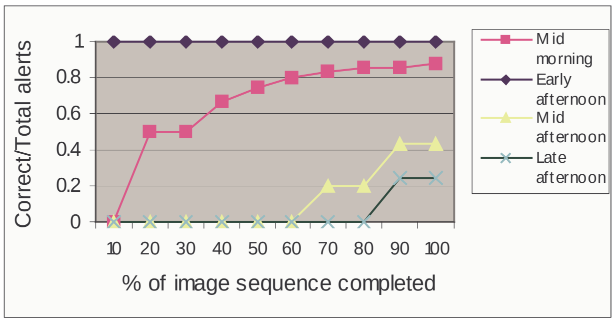
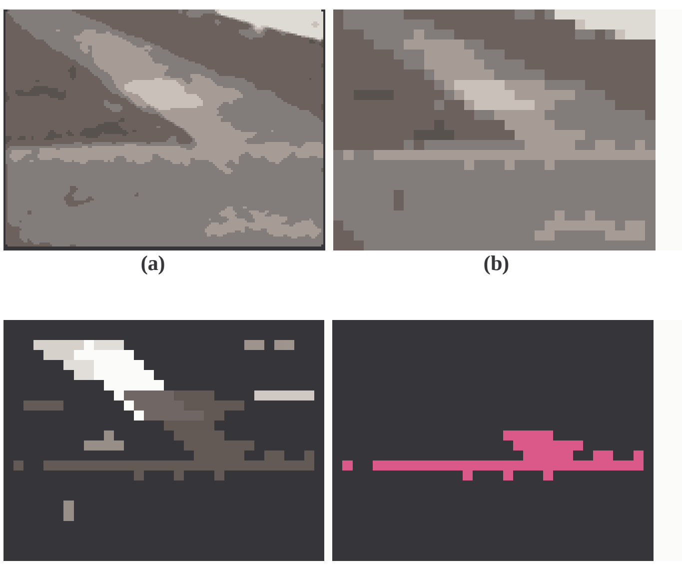


Figure 6. Alert ratios for each bus shelter sequence.



(a) (b) (c) (d)

Figure 7. Example incorrect alert warning.

False alerts were usually shadows on the path, or the area surrounding an obstacle. In figure 7 above, a path shadow is incorrectly identified as an obstacle. The median filtered and 8 grey level image is shown in figure 7a. The 32 x 24 block image (7b) has been segmented in figure 7c. Figure 7d shows the alert segment, which has been incorrectly identified.

CONCLUSION

A functional AHV system needs to provide useful information about the current environment, be reliable, function in near real-time and integrate different visual functions (eg. obstacle avoidance). In this paper a low cost PDA, capable of receiving and processing camera input and outputting simulated phosphenes, has been demonstrated. An original method to simulate AHV and to provide a simple looming obstacle alert has been provided. To improve program efficiency a reduced "block" approximation of each image is used: the use of blocks reduces both memory requirements and the number of calculations required for segmentation and searching for matching segments between successive images. The reduction in grey levels from 8 bits to 3 bits improves performance of the median filter. Ideally each 8 bit 160x120 pixel image would be used for image segmentation and segment matching between images, however limitations in processor and bus speed, limited memory, and the lack of a floating point processor are current technological constraints.

The results of two experiments at four illumination levels have indicated that the initial segmentation and adequate illumination is a significant factor in system performance. The results indicate that the block based method shows promise for development in future AHV systems, although it will be important to consider what ratio of correct alerts versus false alerts will acceptable for system usability.

Future AHV simulation enhancements could utilise colour information: The two obstacles used in this study were both distinctly coloured (red postal box and green bus shelter). Additionally, cheap Global Positioning System (GPS) cards are now available for PDAs and could be integrated to provide useful information on approximate walking speed and location. It should also be possible to utilitise image data (eg. using Bluetooth) from an additional camera, which may allow estimates of depth to be made.

Further experiments are planned with the simulator within an indoor mobility course at QUT. Four different image processing methods will be used to present phosphene simulation displays while participants perform two mobility-related tasks. Results and feedback from these experiments should provide useful information for the future development of the simulation software, and for artificial human vision systems in general.

ACKNOWLEDGMENTS

This research was supported by Cochlear Ltd. and the Australian Research Council through ARC Linkage Grant project 0234229.

Figure 8. Frames 153 (top) to 156 (bottom) of the mid morning post box sequence. The images on the left have been reduced to 8 grey levels and median filtered. On the right is the segmentation result for each image. An obstacle alert (shown in pink) was identified for frame 156.

REFERENCES

1. Dowling, J., *Artificial Human Vision: A review*, Expert Review of Medical Devices. 2005.

2. Cha, K., K. Horch, and R. Normann, *Mobility Performance with a Pixelised Vision System*. Vision Research, 1992. 32(7): p. 1367-1372.

3. Hayes, J.S., et al., *Visually Guided Performance of Simple Tasks Using Simulated Prosthetic Vision*. Artificial Organs, 2003. 27(11): p. 1016-1028.

4. Thompson, R., et al., *Facial recognition using simulated prosthetic pixelized vision.* Investigative Ophthalmolology & Vision Science, 2003. 44(11): p. 5035-5042.

5. Boyle, J.R., A.J. Maeder, and W.W. Boles. *Can Environmental Knowledge Improve Perception with Electronic Visual Prostheses?* Proceedings of the World Congress on Medical Physics and Biomedical Engineering (WC2003). 2003. Sydney, Australia.

6. Gibson, J.J., *The senses considered as perceptual systems*. 1966, Massachusetts: Houghton-Mifflin.

7. Dowling, J., A. Maeder, and W. Boles, *Intelligent image processing constraints for blind mobility facilitated through artificial vision.* Proceedings of the 8th Australian and New Zealand Intelligent Information Systems Conference (ANZIIS), 2003: p. 109-114.

8. Mallot, H.A., *Computational vision : information processing in perception and visual behavior*. 2000, Cambridge, Mass.: MIT Press.