Intelligent Image Processing Constraints for Blind Mobility Facilitated through Artificial Vision

Jason Dowling, Anthony Maeder, Wageeh Boles
School of Electrical and Electronic Systems Engineering, Queensland University of Technology
GPO Box 2434 Brisbane 4001, Queensland Australia
email j.dowling@qut.edu.au   a.maeder@qut.edu.au   w.boles@qut.edu.au

Abstract

This paper examines intelligent image processing constraints which may need to be considered for visual prosthesis development and proposes a display framework which incorporates context, task and alerts related to a scene. A simulation device to investigate this framework is also described. Mobility requirements, assessment, and devices are discussed to identify the functions required by a prosthesis, and an overview of state of the art visual prostheses is provided. Two main computer vision approaches are discussed with application to a visual prosthesis: information reduction and scene understanding.

Keywords

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

1. Introduction

The partial restoration of sight for the blind is an exciting opportunity currently being pursued by a number of international research teams. Partial restoration can be achieved by the use of a camera to capture images, image processing, and then delivery of an electrical impulse to a component of the visual pathway in the human brain. In each case, when this signal is received the subject may perceive multiple points of light ('*phosphenes*'). Because it provides the link between camera and implant, image processing is an integral part of all clinical visual prostheses [1].

Most current mobility devices for the blind provide information in either tactile or auditory form. The two most widely used devices are the *long cane* and *guide dog*. However these devices have limitations: the long cane is only effective over a short range and a guide dog requires expensive training and maintenance. A number of *electronic travel aids (ETA's)* have also been developed over the past thirty years. Early versions of these devices relied on reflected ultrasound to detect obstacles. Later devices have also used laser and image information.

In sighted people, vision usually provides the important functions of mobility (such as avoiding obstacles), therefore it is reasonable to assume that an intelligent image processing (IIP) approach may also provide good mobility information. There are two broad classes of computer vision approaches which could be useful to the development of a visual prosthesis system: information reduction and scene understanding.

The purpose of this paper is to summarise mobility requirements and devices for the vision impaired, before identifying some IIP techniques and constraints that may be important for mobility requirements. A display framework and proposed simulation device are then described.

2. Mobility

Mobility is a person's ability to travel between locations "gracefully, safely, comfortably and independently" [2]. Blind mobility requires skill, effort and training. Mobility problems for a blind person can be caused by changes in terrain and depth (stairs, kerbs); unwanted contacts (bumping); and street crossings (which involve judging the speed and distance of vehicles and may involve identifying traffic light colour) [3]. The most dangerous events for a blind or partially sighted person are drop offs (sudden depth changes, such as on the edge of a subway platform) and moving vehicles [4]. Making unwanted contact with pedestrians is also undesirable as it can be socially awkward and may pose a threat to a persons safety [5].

The assessment of mobility provides information on the effectiveness of different training techniques and also provides information for developers and consumers of mobility devices. The three main methods for assessing mobility are *self report questionnaires*, *field experiments* (such as walking through a shopping mall) and *artificial laboratories* (such as walking through an artificial maze). The latter two methods usually involve a count of the frequency of obstacle contacts by a subject and their walking speed (usually the percentage of preferred walking speed).

3. Mobility information presentation modes

Mobility devices generally provide information by tactile or auditory methods. A visual prosthesis approach is different as it uses a functioning component of the visual pathway, and does not overload another sense. These two presentation modes are summarised in table 1.

Table 1: Summary of blind mobility devices

Presentation Mode	Mobility Device	Sensory method
Conventional	Electronic Travel Aids	Most auditory, some tactile
	Guide dog	Tactile feedback
	Long Cane	Tactile and echolation
	Mobility Robot	Tactile/auditory
	Human Guide	Tactile/auditory
Visual Pathway	Retinal, Cortical and optic nerve prosthesis	Phosphene perception in subject

3.1. Conventional presentation

The most useful devices for mobility are the guide dog and the long cane. However, ambient sound is still important in blind mobility (as it is in sighted mobility). Ambient sound might be directional traffic noise, pedestrian crossing signals or voices. Non-visual information may also be tactile or olfactory.

The long cane provides a blind person with sufficient information for safe movement in the immediate environment. Mobility enhancement in a blind person after long cane training is often dramatic [6]. An additional benefit is the high visibility of the cane to drivers and other pedestrians. The most significant problem with the long cane is that it only provides two paces of preview. Because there is limited preview, a long cane user needs fast reaction times. A cane does not protect against obstacle collision to the upper part of the body (such as wall-mounted public telephones). There is also a risk of tripping other pedestrians with a cane [7].

Guide dogs also provide good mobility assistance by pulling in the same way that a human guide would. They are able to respond to hand and voice signals and are trained to avoid obstacles, prevent veering in street crossings, stop if there is a dangerous situation and intelligently disobey commands that are not safe. A dog may remember common landmarks (such as a particular shop door). However guide dogs are expensive (~AUS$25,000) and require a person to be comfortable with dogs. A guide dog user also needs to be physically fit and prepared to maintain the dog [8].

Although a number of ETA's have been developed, none has achieved widespread market penetration. These devices often provide little benefit in mobility, require a long period of training are expensive and are cosmetically unattractive [6]. Ultrasound based devices (which generally provide auditory information) include the SonicTorch,

SonicGuide and SonicPathfinder. More recent ultrasonic devices include the Navbelt and the Guidecane [9]. A laser cane ETA has also been developed, which uses reflected light energy to provide both tactile and auditory information [10]. The Guide Dogs for the Blind Association of Queensland provide a low cost, handheld ultrasound device, the Miniguide, to visually impaired clients. A promising new ultrasound device, the UltraCane, acts as a standard cane while providing tactile information about head and chest level obstacles in the environment.

The vOICe is the only commercially available ETA which uses an image processing approach. This ETA presents an acoustic representation of images to a blind person [11]. This portable device captures image data using a head mounted camera, processes this information (using a 64x64 pixel array) and provides an auditory representation once per second.

Another approach to blind mobility and navigation involves adjusting the environment to provide useful information. Environmental accessibility for blind and the partially sighted can involve using a logical design layout (for example, stairs should be next to lift), assisting with visibility (for example, hand rails should have high contrast) and adequate lighting (which should be 50-100 % greater than that required for normally sighted) [12]. Useful mobility and navigation information can be provided by locating transmitters at strategic locations in the environment; a blind person can then use a handheld receiver to receive directional information about the landmark. This approach has been used in the 'Talking Signs' program which has been implemented in many locations in San Francisco [13].

3.2. Visual Pathway presentation

Human vision involves the focusing of light by the lens and cornea in the eye and the absorption of this light by photoreceptors in the retina. Electrical signals from these photoreceptors are then processed through a layer of bipolar and ganglion cells within the retina, before passing to the optic nerve [14]. The amount of information entering the eye is reduced considerably: there are over 120 million photoreceptors and only about 1 million ganglion cells. Most signals from the optic nerve pass through the lateral geniculate body to the visual cortex, where a spatial representation of the retinas appears to exist [15].

A visual pathway approach promises to provide significantly more information than a conventional mobility aid, without overloading another sensory channel. The three main approaches to visual prostheses involve a retinal implant (subretinal or epiretinal), an optic nerve cuff electrode and a cortical implant. Each approach has advantages and

disadvantages, and all have been successful in providing phosphene perception to subjects [16]. All methods share a common constraint: due to limitations of current technology there is a limited number of phosphenes that can be generated. Therefore it is necessary to develop methods for the optimal use of these phosphenes.

4. Relevant Intelligent Image Processing techniques

An image processing approach requires an image to be *captured* by a sensor and digitized. This image is then usually *pre-processed* to reduce noise. After these stages we can use an *information reduction* approach to provide essential environmental information, and/or attempt to *understand* objects in the environment. We provide a review of IIP as any visual prosthesis system will need to satisfy the same constraints.

4. 1. Information Reduction (low level vision)

Most existing visual prosthesis efforts are aimed at this level, which is concerned with the reduction or collapse of visual information. Operations on images at this level are designed to improve image saliency, or to emphasize features of particular importance or relevance, for example kerbs or walls. Cortical prosthesis research by the Dobelle Institute has found that *edge detection* and *image reversal* enhance the ability of subjects to recognize important scene components (such as doorways) [1].

Low level computer vision techniques commonly involve image filtering, edge detection and *segmentation* to identify objects within the image. At the information reduction level, a real world scene can be processed and presented in a more visually recognizable form, like a picture or cartoon. It might be sufficient for mobility purposes to have the outline and basic shapes of objects displayed. The reduction of visual information to cartoons, which typically use approximations of features to convey information, can be useful in representing objects using little information [17].

Another low level technique is *motion analysis*. The motion of a person provides visual information about movement relative to the environment and information about the depths of observed scene points [18]. Therefore the analysis of image sequences is desirable in a mobility device. One monocular method of judging depth is *motion parallax* which is used when objects are moving at equal speed: those which are closer to the observer seem to move faster. This information is used for mobility in sighted people (for example, when approaching a subway platform). The detection and location of *moving objects* (such as another person, car or dog) is also important for mobility.

In deciding on the most important features of a scene, it may be useful to consider which parts of a visual scene receive attention from sighted people. Models of the human visual attention have been developed to predict regions of interest in image sequences [19]. Moving objects, people and foreground/centrally located objects have been found to influence subject's attention.

Information reduction ranges from high quality images through to the lowest level of information (binary, low pixel resolution). Previous work at our facility [20], [21], has examined the use of various image processing techniques (such as enhancing edges, using different grey scales and extracting the most important image features) to identify a recognition threshold for low quality stationary images. These images are used to represent the limited number of phosphenes available to the subject (typically a 25x25 array). This research aims at providing a means of determining which parts of a visual scene to represent, and a model for inherent information to determine which visual elements of the scene should be presented for maximum perceptual intelligibility by the subject.

An early visually impaired mobility device which used an information reduction approach was developed during the 1980s. This system [22] used spoken output to describe the current scene. Two modes were provided: the first attempted to identify a safe route for the traveler, using an analysis of gray levels within the image. The second mode used scene analysis in an 'object-identification' mode. This mode attempted to use the aspect ratio of identified objects to categorize an object into three classes: long thin objects (such as a pole or mail box), square or circular objects (such as a pot hole), and large objects (such as a car or wall). When an obstacle was detected, the system provided a warning and asked the user to walk slowly. If object identification was required, the blind person would need to stop walking and wait while this processing took place. An image correspondence technique was used to identify drop-offs.

More recent research [23] has explored the use of edge detection to determine the positions of lines in an image. The grouping of these lines was then used to classify the object (such as a doorway). Paths were also identified using edges and the Hough transform was used to group these into straight lines. A clustering technique, similar to the Hough transform, was then used to find the dominant vanishing point to indicate the subject's direction of travel. A similar approach for a blind mobility device was investigated by the same authors in [24]. This device used stereo vision, combined with sonar for obstacle detection and kerbs. Once an image was captured, edge points were detected and the Hough transform used to locate parallel line

clusters (which were assumed to represent kerb or path information).

4.2 Scene Understanding (high level vision)

This level is concerned with *identifying features* and *extracting information*. The scene structure is still there to a degree, but it is idealised or reduced. An example application might be to identify a bus stop, fire hydrant or traffic light. It may also be useful to know the distance to the object (number of steps, or time at current walking speed).

Due to the limited number of phosphenes that can be generated by current technology, it may be better to present a symbolic representation. For example a small part of the grid (perhaps 5x5) could be used for information on obstacle locations in the current environment. Auditory information could also be provided in natural language. A scene description mode could be useful (similar to [22]).

Object interpretation depends on knowledge of possible objects, and might also depend on context (for example, an outdoor scene versus a home environment. For orientation, it would be useful if a prosthesis system using a scene understanding approach could 'learn' to recognize new objects – for example, an image of a particular type of building (tram stop) could be added to the existing object knowledge base.

Some recent blind mobility research has attempted to recognize particular objects in a scene. The identification of stair cases was addressed in [25]. This research used a texture detection method (using Gabor filters) to locate distant stair cases. Once a person had moved close enough to the stairs, they were then detected by searching for groups of concurrent lines. The intensity variation was then used to partition the convex and concave lines. Homography with some search criteria was then employed to recover the vertical rotation and slope of the located stair case. Although reasonable results were achieved, the approach was found to be slow and not suitable for real-time applications.

5 Visual prosthesis constraints

This section identifies a number of constraints which affect visual prosthesis development:

Number of Phosphenes. Current technology limits the number of phosphenes that can be provided to a patient. Additionally, the size, shape and brightness of phosphenes are not predictable, although recent work in a 4 x 4 retinal prosthesis shows promise in overcoming these problems [26]. The information reduction associated with a 25 x 25 image is demonstrated in figure 2a.

Figure 1a: Original street scene image

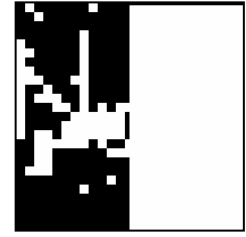**Figure 1b: Street scene with edges detected and binary**

Figure 2a: Subsampled 25 x 25 representation of the street scene in figure 1b.

Figure 2b: The reduced 25 x 25 representation with an 'obstacle on right hand side' alert activated.

Real-time processing: A visual prosthesis system needs to perform in real-time. However, this has been problematic for other image based mobility systems, particularly those that are stereo-vision based. One way of providing real-time processing may be to restrict the field of view of the camera, although this would restrict the amount of preview (or time to anticipate problems) available to a blind traveler. Movement of the subject (egomotion) is an additional image processing overhead, as the camera will need to be attached to the person's body.

Integration/Prioritization : The integration of different functions is a challenge. Dangerous features of the environment, such as moving objects, obstacle detection and sudden changes in depth, should be displayed before less important information. Figure 2b shows a sample obstacle warning in a display, where the right half of the display is highlighted.

Scene type: Scene understanding depends to a large extend on context. An intelligent image processing approach would consider the type of context in recognizing objects. Context information is often defined by previous images, which suggests an image sequence approach. Different display and processing modes will probably be required for a prosthesis system to cope with differences in operational environments.

Device simulation. Currently it is necessary to use a visual prosthesis simulation with normally sighted subjects to test the effects of different IIP approaches on mobility. The results of these investigations may not generalize to actual visual prosthesis patients.

Standard set of image sequences: A constraint on the development of both the information reduction and scene understanding approach for mobility is the lack of a standard set of images/image sequences for evaluation and comparability of methods: an approach which has been successfully applied in the field of Information Retrieval [27]. The benefit of a standard is that different algorithms could be objectively tested against each other and measured for efficiency and effectiveness. Sample image sequences might be captured from a subject walking at normal pace towards different obstacles, over different drop-offs or towards a door. Different environments (such as indoors/outdoors) and lighting conditions could also be included.

6. Proposed mobility display framework

The display from an 'intelligent' visual prosthesis (or simulation) should process different information reduction and scene understanding information depending on the type of scene. For mobility purposes this display depends on three main dimensions (see figure 3) of the current scene:

- *Context:* The type of scene can affect the type of IIP required. For example, there may be a greater need for information reduction in a crowded shopping mall than a suburban street.

- *Task:* Different information is required depending on the current task. A road crossing task may emphasize a straight path to the opposite kerb (to prevent veering), whereas a task involving identifying a set of keys on a cluttered table may involve zooming or object recognition.

- *Alert:* The system needs to continually investigate any hazardous features of the current scene. These alerts, such as an approaching tree branch (obstacle detection) or descending stairs (drop off) need to run as background tasks, and interrupt the current display when required.

Common abstract classes for these dimensions could be developed to reduce decision processing requirements. For example an indoor home and indoor office context could both be assigned to an 'indoor room' class.

7. Simulation system

To investigate the mobility display framework, we are developing a visual prosthesis simulation. This portable head mounted device which consists of a Personal Digital Assistant (PDA) and an attached digital camcorder. Currently the PDA display is used to present the phosphene simulation. A normally sighted subject can then wear the device and be assessed on various mobility tasks under different contexts, alert scenarios and IIP conditions. A sheet of cloth is used to limit the subject's visual information to the PDA display. The proposed simulator components are listed in figure 4.
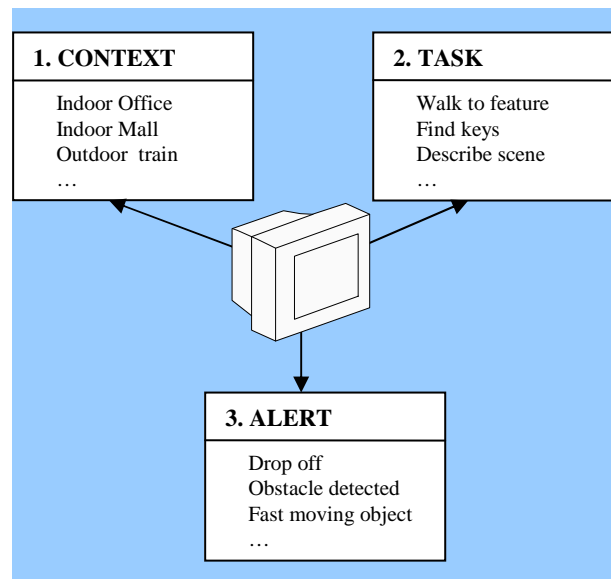

Figure 3: Proposed mobility display framework

Input: PDA digital camera (LifeView FLYCAM-SD)
Processor: Intel XScale PXA255 400 Mhz. (HP iPAQ H5550)
Operating System: Microsoft PocketPC 2003
Simulation display: PDA display (HP iPAQ H5550)
Software Development: MS embedded visual C++, Intel Integrated Performance Primitives

Figure 4: Proposed simulation components

A similar simulation approach was used in [28] where the minimum number of phosphenes required for adequate mobility was investigated. The 'pixelised vision simulator' device consisted of a video camera connected to a monitor in front of the subject's eyes. A perforated mask was placed on the monitor to reproduce the effect of individual phosphenes. The artificial environment consisted of an indoor maze which contained paper column 'obstacles'. Walking speed and frequency of contact were used as dependant variables. This research found that a 25x25 array of phosphenes, with a field of view of 30° would be required for a successful device.

6. Conclusion

A successful visual prosthesis should result in increased mobility performance. Evidence from previous research suggests that the objective assessment of mobility is important in developing and comparing different devices and techniques. In this paper we have described IIP techniques and constraints related to visual prosthesis development. We have also presented a framework and simulation device for further investigation into how IIP techniques can en-

hance the effectiveness of visual prostheses. Experimental work exploring the benefits of this framework is currently under way.

ACKNOWLEDGMENTS

REFERENCES

[1] W. Dobelle, "Artificial Vision for the Blind by Connecting a Television Camera to the Brain," *ASAIO Journal*, vol. 46, pp. 3-9, 2000.

[2] C. A. Shingledecker and E. Foulke, "A human factors approach to the assessment of mobility of blind pedestrians," *Human Factors*, vol. 20, pp. 273-286, 1978.

[3] D. Geruschat and A. J. Smith, "Low vision and mobility," in *Foundations of Orientation and Mobility*, B. B. Blasch and W. R. Weiner, Eds., 2nd ed. New York: American Foundation for the Blind, 1997.

[4] D. G. Pelli, "The visual requirements of mobility," in *Low Vision: Principles and Applications*, G. C. Woo, Ed. New York: Springer-Verlag, 1986, pp. 134-146.

[5] D. Geruschat, K. A. Turano, and J. W. Stahl, "Traditional Measures of Mobility Performance and Retinis Pigmentosa," *Optometry and Vision Science*, vol. 75, pp. 525-537, 1998.

[6] A. Dodds, *Mobility Training for Visually Handicapped People: A Person-Centred Approach*. London: Croom Helm, 1988.

[7] L. W. Farmer and D. L. Smith, "Adaptive technology," in *Foundations of Orientation and Mobility*, B. B. Blasch and W. R. Weiner, Eds., 2nd ed. New York: American Foundation for the Blind, 1997.

[8] R. H. Whitestock, L. Frank, and R. Haneline, "Dog guides," in *Foundations of Orientation and Mobility*, B. B. Blasch and W. R. Weiner, Eds., 2nd ed. New York: American Foundation for the Blind, 1997.

[9] S. Shoval, I. Ulrich, and J. Borenstein, " Computerized Obstacle Avoidance Systems for the Blind and Visually Impaired," in *Intelligent Systems and Technologies in Rehabilitation Engineering*, H. N. L. Teodorescu and L. C. Jain, Eds.: CRC Press, 2000, pp. 414 - 448.

[10] A. Dodds, *Rehabilitating Blind and Visually Impaired People*. London: Chapman and Hall, 1993.

[11] P. B. L. Meijer, "Vision technology for the totally blind," http://www.seeingwithsound.com/, 2003.

[12] B. L. Bentzen, "Environmental accessibility," in *Foundations of Orientation and Mobility*, B. B. Blasch and W. R. Weiner, Eds., 2nd ed. New York: American Foundation for the Blind, 1997.

[13] Talking Signs Inc, "Talking Signs Infrared Communications System," http://www.talkingsigns.com/, 2003.

[14] T. N. Cornsweet, *Visual Perception*. New York: Academic Press, 1970.

[15] R. L. Gregory, *Eye and Brain: The Psychology of Seeing*, 5th ed. Tokyo: Oxford University Press, 1998.

[16] E. Margalit, M. Maia, J. D. Weiland, et al., "Retinal Prosthesis for the Blind," *Survey of Ophthalmology*, vol. 47, pp. 335-356, 2002.

[17] M. C. Fairhrust, *Computer Vision for robotic systems*. New York: Prentice Hall, 1988.

[18] V. S. Nalwa, *A guided tour of computer vision*. Reading, Massachusetts: Addison-Wesley, 1993.

[19] W. Osberger and A. Rohaly, "Automatic Detection of Regions of Interest in Complex Video Sequences," in *Human Vision and Electronic Imaging VI*, vol. 4299. Bellingham, USA: SPIE - The International Society for Optical Engineering, 2001, pp. 361-372.

[20] J. Boyle, A. Maeder, and W. Boles, "Static Image Simulation of Electronic Visual Prosthesis", Proceedings of the 7th Australian and New Zealand Intelligent Information Systems Conference, Perth, pp 85-88, 2001.

[21] J. Boyle, A. Maeder, and W. Boles, "Can environmental knowledge improve perception with electronic visual prostheses?" Proceedings of the 4th World Congress on Medical Physics and Biomedical Engineering, Sydney, 2003.

[22] J. T. Tou and M. Adjouadi, "Computer Vision for the Blind," in *Electronic Spatial Sensing for the Blind*, D. H. Warren and E. R. Strelow, Eds. Dordrecht: Martinus Nijhoff Publishers, 1985, pp. 83-124.

[23] M. Snaith, D. Lee, and P. Probert, "A low-cost system using sparse vision for navigation in the urban environment," *Image and Vision Computing*, vol. 16, pp. 223-292, 1998.

[24] N. Molton, S. Se, M. Brady, D. Lee, and P. Probert, "Robotic Sensing for the Partially Sighted," *Robotics and Autonomous Systems Journal*, vol. 26, pp. 185-201, 1999.

[25] S. Se and M. Brady, "Vision-based Detection of Staircases," Proceedings of the Fourth Asian Conference on Computer Vision ACCV 2000, pp. 535-540, Taipei, January 2000.

[26] M. S. Humayun, J. D. Weiland, G. Y. Fujii, R. Greenberg, R. Williamson, J. Little, B. Mech, V. Cimmarusti, G. Van Boemel, and G. Dagnelie, "Visual perception in a blind subject with a chronic microelectronic retinal prosthesis," *Vision Research*, vol. 43, pp. 2573-2581, 2003.

[27] National Institute of Standards and Technology, "Text Retrieval Conference (TREC)," http://trec.nist.gov/, 2003.

[28] K. Cha, K. Horch, and R. Normann, "Mobility Performance with a Pixelised Vision System," *Vision Research*, vol. 32, pp. 1367-1372, 1992.