

A Substitute Vision System for Providing 3D Perception and GPS Navigation via Electro-Tactile Stimulation

Simon Meers and Koren Ward
School of IT and Computer Science
University of Wollongong
Wollongong, NSW, Australia, 2522
meers@uow.edu.au, koren@uow.edu.au

Abstract

This paper describes a novel approach for enabling the blind to achieve obstacle avoidance and navigation in outdoor environments with the aid of visual sensors, GPS and electro-tactile stimulation. The Electro-Neural Vision System (ENVS) works by calculating a depth map from stereo video cameras by measuring the disparity between the two images. The depth map data is then compressed into a simple range data array and delivered to the fingers via electro-neural stimulation, to indicate to the user the range of objects being viewed by the cameras. To perceive the location of obstacles and the 3D structure of the environment, the user imagines that their hands are held with fingers outstretched in the direction viewed by the cameras, and the intensity of stimulation felt by each finger indicates the distance to objects in the direction pointed at by that finger. Also, the relative location of significant landmarks is determined using GPS and stored GPS coordinates and delivered to the fingers via encoded pulses when the landmarks are in the field of view of the stereo cameras. This intuitive means of perceiving the 3D profile of the environment and the location of landmarks in real-time effectively enables the user to navigate their environment without use of the eyes or other blind aids. Experimental results are provided, demonstrating the potential that this form of 3D environment perception has for enabling the user to achieve localisation, obstacle avoidance and navigation without using the eyes.

Keywords: substitute vision, TENS, electro-tactile, electro-neural vision, stereo cameras, disparity, GPS

1 Introduction

It is estimated there are 37 million people worldwide with severe visual impairment [1]. Consequently, blindness not only represents a severe cognitive handicap, but also a significant burden on the global community. In an attempt to alleviate this problem, we have been investigating the use of electro-tactile user interfaces, stereo video cameras and GPS for providing 3D perception of the environment and landmarks without using the eyes. Our vision system (shown in Figure 1) operates by extracting depth information from the stereo image input and delivering this information to the fingers via electro-neural stimulation. To interpret the range data, the user simply imagines that their hands are held with fingers extended in the direction viewed by the cameras. The amount of electro-neural stimulation felt by each finger indicates the distance to objects in the direction that each finger is pointing.

Furthermore, GPS and stored GPS coordinates are used to indicate to the user the relative location of significant landmarks in the user's vicinity. This is achieved by encoding significant landmarks as a sequence of pulses and delivering the appropriate pulse sequence to the fingers periodically when the corresponding landmark is estimated to be within the

field of view of the stereo cameras. The distance to perceived landmarks is also encoded by the intensity of the pulse sequence.

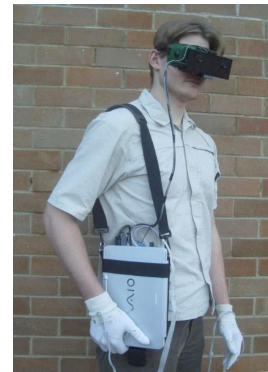


Figure 1. The Electro-Neural Vision System

The continuous delivery of environmental depth information and the location of landmarks to the user, in a form that is easy to interpret, allows the user to realise the 3D profile of their surroundings and the location of known landmarks by surveying the environment with the stereo cameras. This 3D perception can enable the user to navigate the environment, recognise their location and perceive the

size and movement of objects within the surrounding environment without using their eyes.

Section 2 of this paper provides a brief review of previous developments in artificial or substitute vision systems and GPS aids for the blind. Section 3 provides implementation details of the Electro Neural Vision System (ENVS) with respect to 3D perception and landmark recognition. In Section 4 we provide experimental results demonstrating the potential of the ENVS for assisting the blind to navigate outdoor environments. Finally, we provide concluding remarks and briefly outline possible directions for further research and development.

2 Background

Bionic vision in the form of artificial silicon retinas or external cameras that stimulate the retina, optic nerve or visual cortex via tiny implanted electrodes are currently under development (see [2], [3] & [4]). At present, the only commercially available artificial vision implant is the Dobelle Implant [5]. This device provides visual perception comprising a number of points of light. But unfortunately this bears little resemblance to the surrounding environment. Even if more successful results are achieved with implants in the not-so-distant future, many blind people may not have access to implants due to their high cost and the expertise required to surgically implant such devices. Also, some forms of blindness, for example brain or optic nerve damage, may be unsuitable for implants.

A number of wearable devices have been developed for providing the blind with some means of sensing or visualising the environment. For example, Meijer's vOICe [6] compresses a camera image into a coarse 2D array of greyscale values and delivers this information to the ears as a sequence of sounds with varying frequency.

Sonar mobility aids for the blind have been developed by Kay [7]. Kay's system delivers frequency modulated sounds, which represent an object's distance by the pitch of the generated sound and the object's surface texture by the timbre of the sound delivered to the headphones. However, to an inexperienced user, these combined sounds can be confusing and difficult to interpret. Also, the sonar beam from these systems is very specular in that it can be reflected off many surfaces or absorbed resulting in uncertain perception. However, Kay's sonar blind aids can help to identify landmarks by resolving some object features (i.e. resolution, texture, parallax, etc.) and can facilitate some degree of an object's classification to experienced users.

A major disadvantage of auditory substitute vision systems is that they can diminish a blind person's capacity to hear sounds in the environment (e.g. voices, traffic, walking, etc). Consequently, these devices are not widely used in public places because they can reduce a blind person's auditory perception

of the environment and could potentially cause harm or injury if impending danger is not detected by the ears.

Electro-tactile displays for interpreting the shape of images on a computer screen with the fingers, tongue or abdomen have been developed by Kaczmarek *et al* [8]. These displays work by mapping black and white pixels to a matrix of closely-spaced pulsed electrodes which can be felt by the fingers. Although these electro-tactile displays can give a blind user the capacity to recognise the shape of certain objects, such as black alphabetic characters on a white background, they do not provide the user with any useful 3D perception of the environment necessary for navigation, localisation and obstacle avoidance.

In addition to sensing the surrounding environment, it is of considerable benefit if a perception aid can provide the position of the user or nearby landmarks. Currently, a number of GPS devices are available for the blind which can provide the position of the user or specific locations using voice or Braille interfaces. (e.g. Drishti [9], BrailleNote GPS [10], Trekker [11] and others [12], [13]). However, as with audio substitute vision systems, voice interfaces occupy the sense of hearing which can diminish a blind person's capacity to hear important environmental sounds. Braille interfaces are slow, requiring the user to type and read in Braille. Furthermore, these GPS devices provide no means of detecting obstacles.

Our ENVS [14] is significant because it provides a practical, intuitive means of perceiving the 3D structure of the environment in a way that does not reduce a blind person's capacity to hear sounds in the environment. In this paper we describe how the relative location of landmarks in the environment can also be perceived via the ENVS. This allows users to negotiate obstacles in their surroundings while navigating the environment using perceived landmarks. The user can also manually enter landmarks in the ENVS while negotiating the environment so that they can easily navigate to previously visited locations. In the following section we briefly describe the ENVS hardware equipped with GPS and its operation.

3 ENVS Implementation

The basic concept of our ENVS is illustrated in Figure 2. The prototype is comprised of a stereo video camera headset fitted with an electronic compass for obtaining video and direction information from the environment; a laptop computer equipped with a GPS unit for processing the data from the headset and obtaining the GPS location of the user; a Transcutaneous Electro-Neural Stimulation (TENS) unit for generating electrical pulse output which can be felt through the skin; and a pair of gloves containing electrodes to deliver the pulses to nerves in the user's hands.

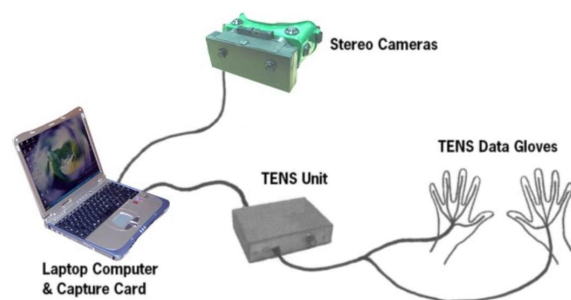


Figure 2. The basic concept of the ENVS

3.1 Obtaining 3D Perception from the ENVS

The ENVS calculates the profile of the surrounding environment by generating a disparity depth map from the images received by the head-mounted stereo video cameras. The depth information is then simplified by taking regional samples which are converted into electrical pulses by the TENS unit which stimulates nerves in the skin via the electrodes in the TENS data gloves (shown in Figure 3). To achieve electrical conductivity between the electrodes and skin, a small amount of conductive gel is applied to the fingers prior to fitting the gloves. For our test purposes, the stereo camera headset is designed to prevent light from entering the user's eyes to simulate blindness.



Figure 3. The ENVS setup

In order for the device to communicate effectively with the user, the signals delivered to the fingers must have a meaningful, intuitive mapping to the data. The depth data sampled from regions spread across the field of vision each correspond to a certain finger; to interpret the signals, the user imagines that their hands are held with fingers extended - the intensity of stimulation felt by each finger is directly proportional to the distance to objects pointed at by that finger.

The software interface shown in Figure 4 allows system parameters (such as disparity settings and TENS output waveform) to be changed during experiments, and allows monitoring and recording of camera, disparity and signal data during use. This figure shows a typical screenshot of the ENVS control panel during operation. The top-left image shows the environment image obtained from one of the cameras in the stereo camera headset. The corresponding

disparity depth map, derived from both cameras, can be seen in the top-right image (lighter pixels have a closer range than darker pixels). Also, the ten disparity map sample regions, used to obtain the ten range readings delivered to the fingers, can be seen spread horizontally across the centre of the disparity map image. These regions are also adjustable via the control panel.



Figure 4. The ENVS control panel.

The intensity of stimulation delivered to each finger is calculated by determining the minimum depth of each of the sample regions. The bar graph, at the bottom-left of Figure 4, shows the actual amount of stimulation delivered to each finger. Using a 1000 MHz Pentium computer we were able to achieve a frame rate of 15 frames per second which proved more than adequate for our experiments.

3.1.1 Extracting Depth from Stereo Video

The depth perception of the ENVS is based on the principle of stereo disparity (see [15]). Just as our eyes capture two slightly different images and our brain combines them with a sense of depth, the stereo cameras in the ENVS captures two images and the laptop computer computes a depth map by estimating the disparity between the two images. However, unlike binocular vision on humans and animals, which have independently moveable eyeballs, typical stereo vision systems use parallel mounted video cameras positioned at a set distance from each other.

3.1.2 Limitations

Calculating the disparity between image pixels the stereo disparity algorithm requires visual features to be detected in the stereo images. Consequently, featureless surfaces can pose a problem for the disparity algorithm due to a lack of identifiable features. For example, Figure 5 illustrates this problem with a disparity map of a whiteboard. As the whiteboard surface has no identifiable features on it, the disparity of this surface and its range cannot be calculated. To make the ENVS user aware of this, the

ENVS maintains a slight signal if a region contains only distant features and no signal at all if the disparity cannot be calculated due to a lack of features in a region. With further work we expect to overcome this deficiency by also incorporating supplementary range sensors (e.g. infrared) into the ENVS headset.

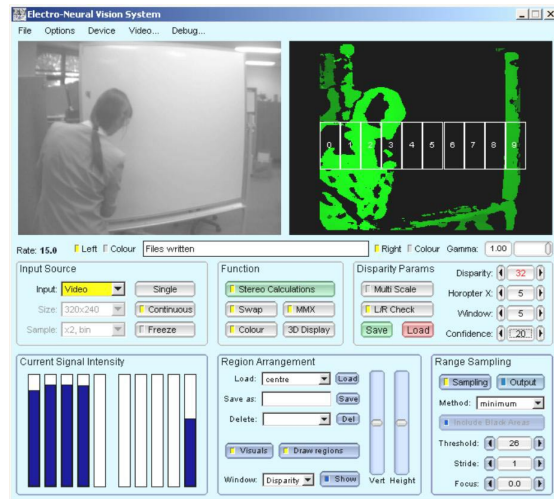


Figure 5. Disparity map of a featureless surface

3.2 Perceiving Landmarks with the ENVS

To enable landmarks to be perceived by blind users, the ENVS is equipped with a GPS unit, an electronic compass and a database of landmarks. The electronic compass, shown in Figure 6, is mounted on the stereo camera headset and is used to determine if the user is looking in the direction of any landmarks.

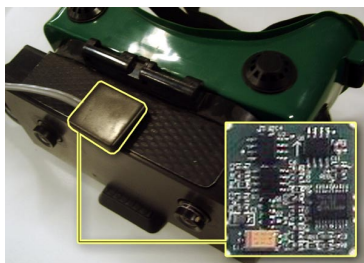


Figure 6. Electronic compass mounted on headset

Landmarks can be loaded from a file or entered by the user by pressing a button on the ENVS and are associated with their GPS location and an ID number. Landmarks are considered to be any significant object or feature in the environment from which the user can approximate their position. A landmark can also be a location the user wants to remember, for example a bus stop or the parking location of a vehicle. By using the GPS unit to obtain the user's location, the ENVS is able to maintain a list of direction vectors to landmarks that are within a set radius from the user. The landmark radius can be set to short or long range (e.g. 200-600m) by the user via a switch on the ENVS unit. When a landmark is calculated to be within the user's ENVS visual field (as determined by the headset compass and the set landmark radius), the ID

of the perceived landmark is encoded into a sequence of pulses and delivered to the user via the finger which represents the direction of the landmark. For example, if a landmark is determined to be in the far left visual field, the pulse sequence corresponding to the landmark will be felt on the left pinky finger.

To encode the ID of a landmark, a 5-bit sequence of dots and dashes carried by a 200Hz signal is used to represent binary numbers. To avoid interfering with the range readings of objects, which are also delivered to the fingers via the ENVS (see Section 3.1), locations are delivered to the fingers in five-second intervals. Thus if a landmark is detected, the user will receive range readings via the fingers for four seconds followed by approximately one second of landmark ID information. If more than one landmark is present within the set landmark radius and the visual field of the cameras, the landmark nearest to the centre of the visual field will be output to the user.

By using five bits to represent landmark IDs, the ENVS is able to store up to thirty-two locations, which proved adequate for our experiments. The distance to the landmark is indicated by the intensity of the pulses. Weak sensations indicate that the landmark is near to the set maximum landmark radius. Strong sensations indicate that the landmark is only metres away from the user. If the user has difficulty recognising landmarks by their pulse sequence, a button is available on the ENVS unit to output the location, distance and bearing of the perceived landmark as speech.

4 Experimental Results

To test the ENVS we conducted a number of experiments within the campus grounds with different users, to determine the competency with which the users could navigate the campus environment without any use of their eyes. All of the ENVS users were familiar with the campus grounds and the landmarks stored in the ENVS and had no visual impairments. To simulate blindness with sighted users the stereo camera headset was designed to fit over the user's eyes so that no light whatsoever could enter the eyes. All reported tests were conducted on users who had at least one hour practise at using the ENVS.

4.1 Navigating the Car Park

Our first test was performed to determine if the ENVS users could navigate a car park and arrive at a target vehicle location that was encoded into the ENVS as a landmark. Each user was fitted with the ENVS and lead blindfolded to a location in the car park that was unknown to them and asked to navigate to the target vehicle using only the ENVS electro-tactile signals. The car park was occupied to approximately 75% of its full capacity and also contained some obstacles such as green strips and lampposts.

Generally we found all users were able to perceive and describe their surroundings and the location of the target vehicle in sufficient detail for them to be able to navigate to the target vehicle without bumping into cars or lampposts. We found that experienced users could also interpret the ENVS output without needing to extend their hands to assist with visualisation and could walk between closely spaced vehicles without colliding with the vehicles.

Figure 7a shows a user observing two closely spaced vehicles in the car park with the ENVS. The profile of the space between the vehicles can be seen in the disparity map, shown in the top right of Figure 7b, and in the finger pulse histogram shown at the lower left of Figure 7b. The yellow bar at the left forefinger position of the finger histogram indicates that target vehicle is located slightly to the left of where the user is looking and at a distance of approximately 120m.



(a)



(b)

Figure 7. (a) Photo of an ENVS user negotiating the car park. (b) The ENVS screenshot showing the perceived vehicles and the target landmark's direction and distance.

4.2 Navigating the Campus

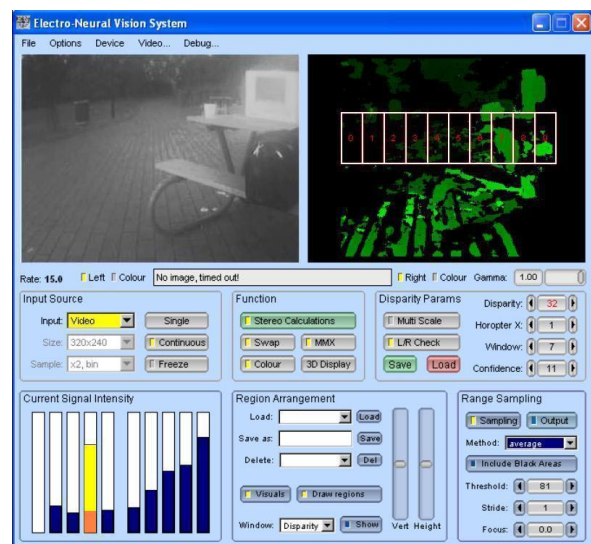
Experiments were also performed within the campus to determine the potential of the ENVS to enable

blindfolded users to navigate the campus without using other aids. The main test was to see if users could navigate between two locations some distance apart (approx. 500m) and avoid any obstacles that might be in the way. All users who performed this test were familiar with the campus environment and had some experience at using the ENVS. The path was flat and contained no stairs between the two locations. A number of familiar landmarks were stored in the ENVS in the vicinity of the two locations.

We found that all users were able to avoid obstacles, report their approximate location and orientation, and arrive at the destination without difficulty. Unlike the car park, the paved path was highly textured making it clearly visible to the disparity cameras. Consequently, this delivered clearly defined range readings of the paved path to the user via the ENVS unit as shown in Figure 8.



(a)



(b)

Figure 8. (a) Photo of an ENVS user surveying a paved path in the campus environment. (b) The associated ENVS screenshot.

This presented a minor problem for some inexperienced users as they were unable to determine where the path ended and the grass began in some places using the ENVS range stimulus alone. However, this did not cause any collisions and the

users became quickly aware of the edge of the path whenever their feet made contact with the grass. We hope to overcome this minor problem by encoding colour into the range signals delivered to the fingers by varying the frequency of the tactile signals.

5 Conclusion

This paper describes our preliminary experimental results in developing a substitute vision system for the blind capable of providing 3D perception of the environment and landmark identification via an electro-tactile interface. Existing substitute vision systems do not provide output in a form which allows intuitive derivation of three-dimensional environment information. Also, existing substitute vision systems do not provide adequate means for the user to perceive landmarks that are important for navigation. Although commercial GPS units are available for the blind, they generally use an audio or voice interface which can interfere with a blind person's ability to hear important environmental sounds. The sounds made by talking GPS units can also undesirably attract the attention of nearby individuals.

Our outdoor experimental results demonstrate that the ENVS can provide a blindfolded user with the ability to avoid obstacles, navigate the environment and locate their position in the environment using familiar landmarks by interpreting sensory data via the electro-tactile data gloves.

With further research and development we hope to develop the ENVS into an effective device capable of providing the blind with increased environment perception and autonomy. Possible future research directions include the provision of colour perception, advanced processing of sensor data to detect and recognise objects, multimodal output, development of compact hardware, fabrication of alternative TENS user interfaces, and use as a computer interface.

6 Acknowledgements

This work was undertaken with the assistance of an Australian Research Council Discovery Grant.

7 References

- [1] World Health Organisation, Magnitude and Causes of Visual Impairment, *Fact Sheet No.282*, November 2004.
- [2] Wyatt, J.L. and Rizzo, J.F., Ocular Implants for the Blind, *IEEE Spectrum*, Vol.33, pp.47-53, May 1996.
- [3] Rizzo, J.F. and Wyatt, J.L., Prospects for a Visual Prosthesis, *Neuroscientist*, Vol.3, pp.251-262, July 1997.
- [4] Rizzo, J.F. and Wyatt, J.L., Retinal Prosthesis, in: *Age-Related Macular Degeneration*, J.W. Berger, S.L. Fine and M.G. Maguire, eds., Mosby, St. Louis, pp. 413 - 432. 1998.
- [5] Dobelle, W. Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex, *American Society of Artificial Internal Organs Journal*, January/February 2000.
- [6] Meijer, P.B.L. An Experimental System for Auditory Image Representations, *IEEE Transactions on Biomedical Engineering*, Vol. 39, No. 2, pp. 112-121, February 1992.
- [7] Kay, L. Auditory Perception of Objects by Blind Persons Using Bioacoustic High Resolution Air Sonar. *JASA*, Vol 107, pp 3266-3275, No 6, June 2000.
- [8] Kaczmarek, K.A. and Bach-y-Rita, P., *Tactile Displays, in Virtual Environments and Advanced Interface Design*, Barfield, W. and Furness, T., Eds. New York: Oxford University Press, pp. 349-414, 1995.
- [9] Helal, A.; Moore, S.E.; Ramachandran, B., Drishti: an integrated navigation system for visually impaired and disabled, *Proceedings, Fifth International Symposium on Wearable Computers*, 2001.
- [10] Mosen, J.; What's New with Pulse Data International's BrailleNote PDA for the Blind, *Center on Disabilities Technology and Persons with Disabilities Conference*, 2005.
- [11] Daviault, J.; Study On The Benefits Of GPS-Based Orientation Tools for the Mobility and Orientation of the Blind, *Center on Disabilities Technology and Persons with Disabilities Conference*, 2005.
- [12] Balachandran, W.; Cecelja, F.; Ptasinski, P., A GPS based navigation aid for the blind, *ICECom*, 2003
- [13] Virkes, D.; Nenadic, K.; Virkes, D.Z., Positioning systems for mobility of the blind, overview and analysis, *Proceedings, 46th International Symposium Electronics in Marine (ELMAR)*, 2004.
- [14] Meers, S. and Ward, K., A Vision System for Providing 3D Perception of the Environment via Transcutaneous Electro-Neural Stimulation, *Proceedings of the 8th IEEE International Conference on Information Visualisation*, London, pp.546-552, July 2004.
- [15] Banks, J. Bennamoun, M. and Corke, P., Non-Parametric Techniques for Fast and Robust Stereo Matching, *Proceedings of the IEEE TENCON'97*, Brisbane, December 1997.