

# A Vision System for Providing 3D Perception of the Environment via Transcutaneous Electro-Neural Stimulation

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

## Abstract

*The development of effective user interfaces, appropriate sensors, and information processing techniques for enabling the blind to achieve additional perception of the environment is a relentless challenge confronting HCI and sensor researchers. To address this challenge we have developed a novel 3D vision system that can enable the 3D structure of the immediate environment to be perceived via head mounted stereo video cameras and electro-tactile data gloves without requiring any use of the eyes. The electro-neural vision system (ENVS) works by extracting a depth map from the camera images by measuring the disparity between the stereo images. This range data is then delivered to the fingers via electro-neural stimulation to indicate to the user the range of objects being viewed by the cameras. To interpret this information, the user only has to imagine that the hands are held in the direction viewed by the cameras, with fingers extended, and the amount of stimulation felt by each finger indicates the range of objects in the direction pointed at by each finger. This intuitive means of perceiving the 3D structure of the environment in real time effectively enables the user to navigate the 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 at enabling the user to achieve localisation and obstacle avoidance skills without using the eyes.*

**Keywords---** substitute vision, TENS, electro-tactile, electro-neural vision, stereo cameras, disparity.

## 1. Introduction

It is difficult to imagine something more profoundly disabling than losing the sense of sight. Yet blindness occurs to many thousands of people every year as a result of injury, disease or birth defects. To address this problem, we have been experimenting with electro-tactile user

interfaces and stereo video cameras for providing the user with useful 3D perception of the environment without using the eyes. Our vision system works by extracting depth information from the stereo cameras and delivering this information to the fingers via electro-neural stimulation. To interpret the range data, the user only has to imagine that the hands are being 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 of each of the fingers, as shown in Figure 1.



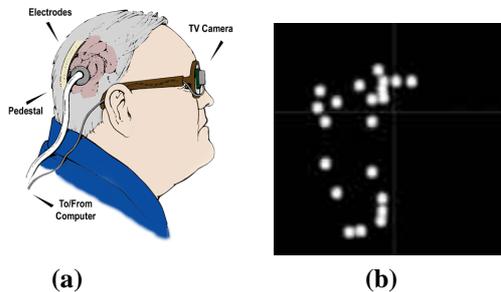
**Figure 1. The Electro-Neural Vision System**

By having environmental depth information delivered continuously to the user in a form that is easy to interpret, the user is able to realise the 3D profile of the environment and the location of objects in the environment by surveying the environment with the cameras. This form of 3D environment perception can then be used to navigate the environment, recognise the user's location in the environment and perceive the size and movement of objects within the environment without using the eyes.

In Section 2 of this paper we provide a brief review of previous work done on artificial or substitute vision systems for the blind. In Section 3 we provide details of the user interface and operation of the Electro Neural Vision System (ENVS). Section 4 discusses the basic theory and limitations of extracting depth information from the environment with stereo cameras. In Section 5 we provide the results of experiments we have conducted in our laboratory with the ENVS. Finally, we provide concluding remarks and a brief description of further work to be done.

## 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 [1], [2] & [3]). Currently, the only commercially available artificial vision implant is the Dobbelle Implant [4]. This is comprised of an external video camera connected to a visual cortex implant via a cable, as shown in Figure 2(a). Once implanted, this provides the user with visual perception in the form of a number of perceivable “phosphenes”, as shown in Figure 2(b). Unfortunately, this form of perception bears no resemblance to the environment and has only been demonstrated to be useful for simple classification tasks like learning to classify a small set of large alphabetic characters.



**Figure 2. The Dobbelle Brain Implant. (a) The visual cortex implant. (b) The resulting available vision.**

Even if more successful results are achieved with implants in the not so distant future, many blind people may not benefit from implants due to their high cost and the expertise required to surgically implant the device. Some forms of blindness (eg. brain or optic nerve damage) may also be unsuitable for implants.

In addition to bionic vision implants, a number of wearable devices are either available or under development for providing the blind with some means of sensing or visualizing the environment. One such device, developed by Meijer [5] and named the vOICe, attempts to provide the user with visual cognition by encoding camera image data into sounds, (see Figure 3(a)). This is done by compressing the camera image into a coarse 2D array of grayscale values, as shown in Figure 3(b), and by then

converting each grayscale element into a sound with a specific frequency. This audio information is then delivered to the ears via headphones by sequentially scanning the 2D array of sounds row by row until the entire soundscape is heard.



**Figure 3. (a) The vOICe auditory substitute vision system. (b) A soundscape image.**

However, there are no reported tests done with the vOICe indicating any increased obstacle avoidance or navigational skills from this form of auditory visual perception. It appears, there simply is too much information comprising video frames for any significant auditory interpretation to be possible by this means in real time. Even if it were possible for a user to mentally reconstruct an image’s original greyscale grid by carefully listening to the image’s soundscape, this grid would be either too coarse to reveal any environmental details, or would take too long to listen to for real time cognitive image processing to be possible. Furthermore, by being a coarse 2D greyscale representation of a 3D environment, it may also be impossible for the user to perceive the locations of objects in 3D space which is necessary for obstacle avoidance and navigation. Consequently, little benefit is able to be demonstrated by users wearing this device apart from doing some simple tasks like identifying the direction of an isolated linear object or finding a significant object lying on a uniformly coloured floor.

Considerable work on sonar mobility aids for the blind has been done by Kay [6]. Kay’s work is significant because his Binaural, Trisensor and Sonic Torch sonar systems (see Figure 4) utilise frequency modulated signals, 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.

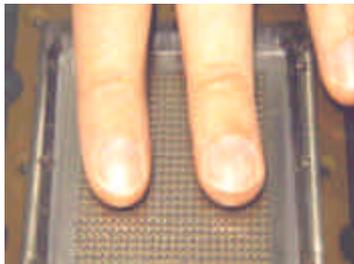


**Figure 4. Kay’s Sonic Torch.**

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. Nevertheless, Kay's device can help to identify landmarks by resolving some object features (i.e. resolution, texture, parallax) that can facilitate some degree of object classification to experienced users.

A further drawback of auditory substitute vision systems is that by using the ears as their information receptor, they can diminish a blind person's capacity to hear sounds in the environment, (eg voices, traffic, walking, etc). Consequently, these devices are not widely used in public places because they can actually reduce a blind person's perception of the environment and could potentially cause harm or injury by reducing a blind person's capacity to detect impending danger from sounds or noise, (eg moving cars, people calling out, alarms, a dog barking, etc).

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* [7], (see Figure 5.) These displays work by simply 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, like black alphabetic characters on a white background, they do not provide the user with any useful 3D perception of the environment which is needed for environment navigation, localization, landmark recognition and obstacle avoidance.

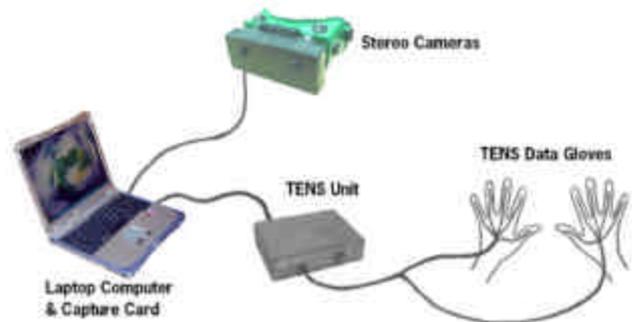


**Figure 5. Kaczmarek's electro-tactile display.**

Our ENVS is significant, not only because it does not impede a blind person's capacity to hear sounds in the environment, but because it provides a useful intuitive means of perceiving the 3D location of objects within the environment. This makes it possible for a user to navigate the environment while avoiding obstacles. The user can also realise his or her location within the environment by perceiving the 3D profile of the environment and by recognising where significant objects are located within this 3D space. In the following section, we provide a brief description of the ENVS setup, operation and user interface.

### 3. The ENVS User Interface and Operation

The basic concept of the ENVS is shown in Figure 6. The ENVS is comprised of a stereo video camera headset for obtaining video information from the environment, a laptop computer for processing the video data, a Transcutaneous Electro-Neural Stimulation (TENS) unit for converting the output from the computer into appropriate electrical pulses that can be felt via the skin, and special gloves fitted with electrodes for delivering the electrical pulses to the fingers.



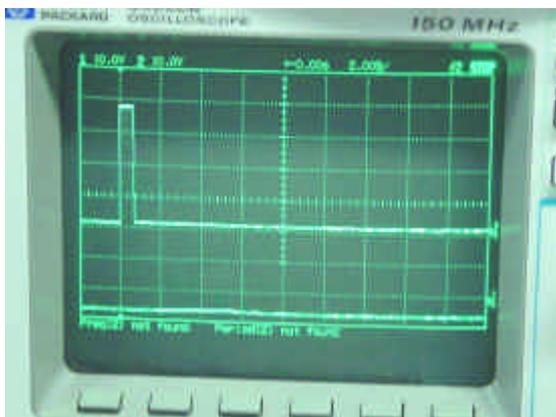
**Figure 6. The basic concept of the ENVS**

The ENVS works by using the laptop computer to obtain a disparity depth map of the immediate environment from the head mounted stereo cameras. This is then converted into electrical pulses by the TENS unit that stimulates nerves in the skin via electrodes located in the TENS data gloves. 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 testing purposes, the stereo camera headset is designed to completely block out the users eyes to simulate blindness. Our ENVS setup is shown in Figure 7.



**Figure 7. The ENVS setup**

The key to obtaining useful environmental information from the electro-neural data gloves lies in representing the range data delivered to the fingers in an intuitive manner. To interpret this information the user imagines his or her hands are positioned in front of the abdomen with fingers extended. The amount of stimulation felt by each finger is directly proportional to the distance of objects in the direction pointed by each finger. Figure 8 shows an oscilloscope screen shot of a typical TENS pulse. To conduct our experiments we set the TENS pulse frequency to 20 Hz and the amplitude to between 40V to 80V depending on individual user comfort. To control the intensity felt by each finger the ENVS adjusts the pulse width between 10 to 100 $\mu$ s.

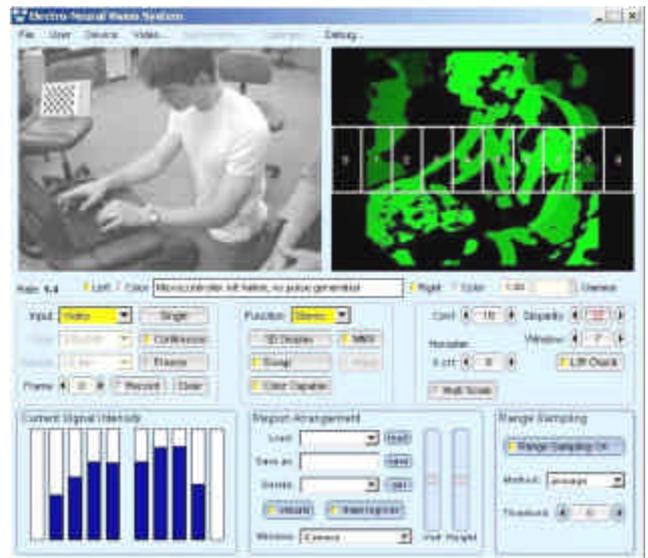


**Figure 8. The TENS output waveform**

We found adjusting the signal intensity by varying the pulse width preferable to varying the pulse amplitude for two reasons. (1) It enabled the overall intensity of the electro-neural simulation to be easily set to a comfortable level by presetting the pulse amplitude. (2) It also simplified the TENS hardware considerably by not needing any digital to analogue converters or analogue output drivers on the output circuits.

To enable the stereo disparity algorithm parameters and the TENS output waveform to be altered for experimental purposes we provided the ENVS with the control panel shown in Figure 9. This was also designed to monitor the image data coming from the cameras and the signals being delivered to the fingers via the TENS unit.

Figure 9 shows a typical screen grab of the ENVS's control panel while in operation. The top-left image shows a typical 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 (i.e. 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 9. The ENVS control panel.**

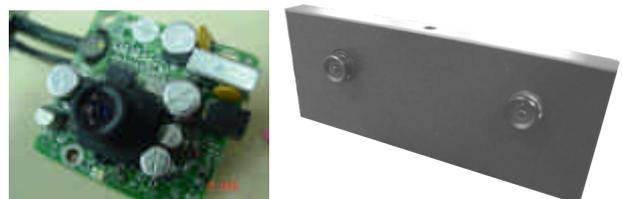
To calculate the amount of stimulation delivered to each finger, the minimum depth of each of the ten sample regions is taken. The bar graph, at the bottom-left of Figure 9, shows the actual amount of stimulation delivered to each finger. Using a 450 MHz Pentium 3 computer we were able to achieve a frame rate of 15 frames per second which proved more than adequate for our experiments.

## 4. Extracting Depth Data from Stereo Video

The ENVS works by using the principle of stereo disparity. 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 eye balls, typical stereo vision systems use parallel mounted video cameras positioned at a set distance from each other.

### 4.1. The Stereo Camera Head

For our experimentation we have been using a pair of parallel mounted DCAM video cameras manufactured by Videre Design [8], as shown in Figure 10. The stereo DCAMs interface with the computer via the firewire port.



**Figure 10. The Stereo DCAMs**

## 4.2. Calculating Disparity

The process of calculating a depth map from a pair of images using parallel mounted stereo cameras is well known [9]. By knowing the baseline distance between the two cameras and their focal lengths (shown in Figure 11), the coordinates of corresponding pixels in the two images can be used to derive the distance to the object from the cameras at that point in the images.

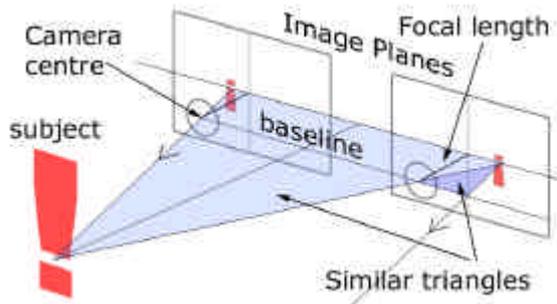


Figure 11. Stereo disparity geometry

Calculating the disparity between two images involves finding corresponding features in both images and measuring their displacement on the projected image planes. For example, given the camera setup shown in Figure 11, the distance from the cameras to the subject can be calculated quite simply. If we let the horizontal offsets of the pixel in question from the centre of the image planes be  $x_l$  and  $x_r$  for the left and right images respectively and the focal length be  $f$  with the baseline  $b$ . By using the properties of the similar triangles denoted in Figure 11, then  $z = f(b/d)$ , where  $z$  is the distance to the subject and  $d$  is the disparity ( $x_l - x_r$ ). To compute a complete depth map of the observed image in real time is also computationally expensive because the detection of corresponding features and calculating their disparity has to be done at frame rate for every pixel on each frame.

## 4.3. Limitations

The stereo disparity algorithm requires automated detection of corresponding pixels in the two images, using feature recognition techniques, in order to calculate the disparity between the pixels. Consequently, featureless surfaces can pose a problem for the disparity algorithm due to a lack of identifiable features. For example, Figure 12 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 ENVIS user aware of this, the ENVIS 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. We expect to overcome this deficiency by also incorporating IR range sensors into the ENVIS headset.

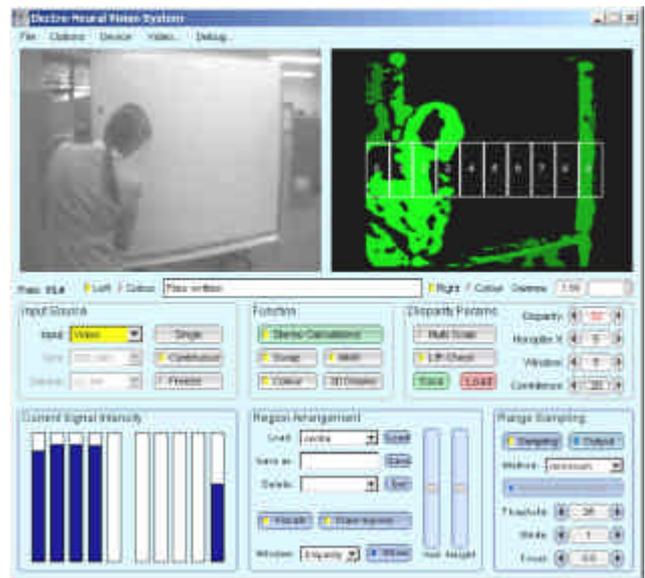


Figure 12. Disparity map of a featureless surface

## 5. Experimental Results

To test the ENVIS we conducted a number of experiments with different users to determine the extent to which the users could navigate our laboratory environment and recognize his or her location within this environment without any use of the eyes. At this point in time we have not conducted experiments with blind users. To simulate blindness with sighted users the stereo camera headset was designed and fitted over the user's eyes so that no light whatsoever could enter the eyes. All reported tests were conducted on five users who had less than 1 hour prior experience using the ENVIS.

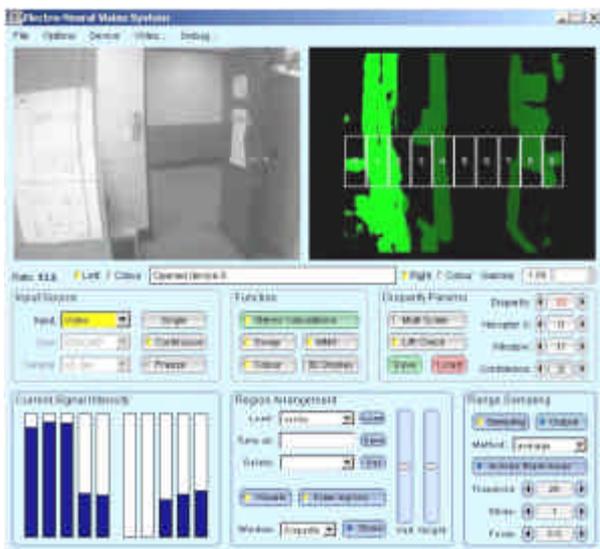
### 5.1. Obstacle Avoidance

Our first tests were done mainly to find out if the user could identify and negotiate obstacles while moving around in the cluttered laboratory environment. We found after 5 minutes of use within the unknown environment, all the users could estimate the direction and range of obstacles located in the environment with sufficient accuracy for the user to be able to approach objects and then walk around them by interpreting the range data delivered to the fingers via the ENVIS. As our environment contained many different sized obstacles, it was also necessary for users to regularly scan the immediate environment, (mostly with up and down head movements), to ensure all objects were detected regardless of their size. After 10 minutes moving around in the environment, while avoiding obstacles, we found it was possible for most users to also identify features like the open doorway, shown in Figure 13 and even walk through the doorway by observing this region of the environment with the stereo cameras. Figure 13 shows a photo of a user and a screen dump of

the ENVS control panel at one instant while the user was performing this task. The 3D profile of the doorway can be plainly seen in the depth map shown at the top right of Figure 13(b). Also, the corresponding intensity of the TENS pulses felt by each finger can be seen on the bar graphs shown at the bottom left corner of Figure 13(b). Although 10 range readings delivered to the fingers this way may not seem like much environmental information, the real power of the ENVS is due to the user being able to easily interpret the 10 range readings and by fusing this information over time, produce a mental 3D picture of the environment. Remembering the locations of obstacles was found to increase with continued use of the ENVS eliminating much of the need to regularly scan the environment comprehensively. Instead experienced users would tend to only use the cameras to confirm the known existence and location of objects. Experienced users could also interpret the range data without any need to hold the hand in front of the abdomen.



(a)



(b)

**Figure 13.** Photo and ENVS screen dump of the user surveying a doorway. (a) The doorway (b) Screen dump of the ENVS

## 5.2. Localization

We also conducted localization experiments to determine if the user could recognize his or her location within the laboratory environment after becoming disoriented. This was performed by rotating the user a number of times on a swivel chair, in different directions, while moving the chair. Care was also taken to eliminate all noises in the environment that might enable the user to recognize the locations of familiar sounds. We found that as long as the environment had significant identifiable objects that were left unaltered and the user had previously acquired a mental 3D map of the environment, the user could recognize significant objects, recall his/her mental map of the 3D environment and describe approximately where he/she was located in the environment after surveying the environment for a few seconds. However, this task becomes more difficult if the environment lacks significant perceivable features or is symmetrical in shape.



(a)



(b)

**Figure 14.** Photo and ENVS screen dump of the user while surveying the environment. (a) The environment. (b) ENVS screen dump.

Figure 14 shows a photo of a user and a screen dump of the ENVIS control panel at one instant while a user was surveying the environment to determine his location. The approximated height, width and range of the table in the foreground of Figure 14(a) can be plainly seen in the depth map, shown at the top right of Figure 14(b). The corresponding intensity of the TENS pulses felt by each finger can be seen on the bar graphs shown at the bottom left corner of Figure 14(b).

The inability of stereo cameras to resolve the depth of featureless surfaces was not a problem within our cluttered laboratory environment because there were sufficient edges and features on the lab's objects and walls for the disparity to be resolved from the stereo video images. In fact, not resolving the depth of the floor benefited our experiments to some extent by enabling objects located on the floor to be more clearly identifiable, as can be seen in Figure 14. However, as explained in Section 4.3, the inability of stereo cameras to resolve the range of featureless surfaces can pose a serious problem for the user in environments that contain flat featureless walls and/or large objects. To overcome this problem we intend incorporating infrared range sensors or beam projectors into the stereo head to enable the range of such surfaces to be resolved.

## 6. Conclusion

The main problem with existing attempts at providing the blind with artificial vision is that the information delivered to the user is in a form that is either hard for the user to understand or difficult for the brain to derive a 3D model of the environment from. Consequently, most existing artificial vision systems, intended for use by the blind, do not adequately provide the 3D perception necessary for avoiding obstacles or navigating the environment. To address this deficiency we have developed an Electro Neural Vision System (ENVIS) based on extracting range data from the environment that is delivered to the user via electro-tactile stimulation in a manner that enables the user to perceive the 3D structure of the environment.

Our preliminary experimental results demonstrate that our ENVIS is able to provide the user with the ability to avoid obstacles, navigate the environment and locate his or her position within our laboratory environment without any use of the eyes. With further work we hope to develop

the ENVIS into an effective device capable of providing the blind with increased environment perception and autonomy. This additional work includes the incorporation of infrared range sensors into the headset for detecting the range of featureless surfaces, the use of pulse coded electro-tactile stimulation for identifying certain colours or landmarks, the development of compact hardware for reducing the bulkiness of the ENVIS and the fabrication of alternative TENS garments eliminating the need for the user to wear gloves.

## Acknowledgements

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

## References

- [1] Wyatt, J.L. and Rizzo, J.F., Ocular Implants for the Blind, *IEEE Spectrum*, Vol.33, pp.47-53, May 1996.
- [2] Rizzo, J.F. and Wyatt, J.L., Prospects for a Visual Prosthesis, *Neuroscientist*, Vol.3, pp.251-262, July 1997.
- [3] 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.
- [4] Dobbelle, 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.
- [5] Meijer, P.B.L. An Experimental System for Auditory Image Representations, *IEEE Transactions on Biomedical Engineering*, Vol. 39, No. 2, pp. 112-121, Feb 1992. Reprinted in the *1993 IMIA Yearbook of Medical Informatics*, pp. 291-300.
- [6] 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.
- [7] 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.
- [8] Videre Design url: <http://www.videredesign.com>
- [9] Banks, J. Bennamoun, M. and Corke, P., Non-Parametric Techniques for Fast and Robust Stereo Matching. In *IEEE TENCON'97*, Brisbane, Australia, December 1997.