34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Smart Image Processing System for Retinal Prosthesis

James D. Weiland, Neha Parikh, Vivek Pradeep, Gerard Medioni

Abstract— **Retinal prostheses for the blind have demonstrated the ability to provide the sensation of light in otherwise blind individuals. However, visual task performance in these patients remains poor relative to someone with normal vision. Computer vision algorithms for navigation and object detection were evaluated for their ability to improve task performance. Blind subjects navigating a mobility course had fewer collisions when using a wearable camera system that guided them on a safe path. Subjects using a retinal prosthesis simulator could locate objects more quickly when an object detection algorithm assisted them. Computer vision algorithms can assist retinal prosthesis patients and low-vision patients in general.**

I. Introduction

A retinal prosthesis uses electrical stimulation to restore the sensation of vision, in people with retinal degenerative disease.[1] A microelectrode array for patterned stimulation of the retina is positioned either on the ganglion cell side of the retina (epi-retinal) or in the place vacated by the degenerated photoreceptors (subretinal). Sensory substitution devices are also being tested, that convert images into tactile or aural patterns, which the blind person feels or hears, then interprets this sensation to understand the image.[2],[3] Both prosthetic and substitution devices have an external camera and video processing hardware to digitize images.

Clinical trials of prototype retinal prostheses have produced encouraging results. Even with only 60 pixels, implant patients with the Argus II retinal prosthesis (Second Sight Medical Products, Inc., Sylmar, CA, USA) can recognize large letters and have improved mobility.[4] Other studies have shown similar, encouraging results.[5] However, the visual perceptions created by retinal implants are still crude relative to natural vision. Thus, it is important to consider how to improve the overall experience and abilities of these subjects through image processing.

Image processing for a retinal prosthesis can take two forms. One is the real-time conversion of video into a stimulus pattern that is then applied to the retina via the implanted stimulator. Optimization of this conversion process to achieve the best visual performance is critical to the success of retinal prosthesis. A second type of image processing, which is the topic of this report, is using computer vision algorithms to interpret a scene and provide this interpretation to the user, in the form of visual, aural, or tactile cues. Visual information could be communicated through the electrode array, aural information via headphones, and tactile information via vibration motors. Herein, we report on pilot studies of two prototype systems which used computer vision algorithms with guiding cues for navigation and object detection.

Research is supported by Telemedicine & Advanced Technology Research Center (TATRC), at Fort Detrick, MD under Contract Number: W81XWH-10-2-0076 and by the National Science Foundation under Grant No. EEC-0310723. J.D. Weiland is with the Doheny Eye Institute/University of Southern California. N. Parikh was with the University of Southern California and is now with Medtronic, Inc. V. Pradeep was with the University of Southern California and is now with Microsoft, Inc. G. Medioni is with the University of Southern California.

II. Methods

The primary components of navigation system are shown in Figure 1. A similar prototype has been reported previously.[6] The stereo camera (Tyzx, Deep Sea V2) generates dense 3D maps of the scene similar to a range-finder. This 3D data is used as input to algorithms that perform obstacle detection and compute the shortest and safest path through the environment. However, as the camera system is mounted on the head, it undergoes significant pitch and yaw motions during walking. Furthermore, to guide a human subject through the environment, the system must have knowledge about the location and orientation of the user's head with respect to the obstacle map. The user's motion should be continually tracked (localization) on a global frame of reference, and new obstacles be registered into this frame (mapping). The algorithm performs all of these functions. An advantage of this simultaneous localization and mapping is that the system has memory about obstacles present in the vicinity, even when the objects are not within the view of the camera.

If the subject deviates from the safe path (which is computed by locating the intersection between direction of motion and a line segment on the shortest path), corresponding vibration motors (left/right shoulder) are activated for cuing a corrective rotation. As a failsafe, if an obstacle is present in close proximity (within 1 m), the waist motors (right/left) are activated to cue the user to step away from it.

To test navigation system, 9 subjects with varying visual impairment conditions were recruited at the Braille Institute of Los Angeles. Subjects were instructed on basic operation of the system and the vibration motors were manually activated to allow subjects to become familiar with the sensation created by the motors. An obstacle course was set up (figure 1) that the subjects navigated through 10 times. Subjects were randomly divided into 3 groups: group 1 used the white cane, group 2 used the device and group 3 used the cane and device simultaneously. After a month, group 1 and group 2 subjects were swapped to ensure that results were not biased because of group composition.

Figure 1 – Top - Overhead view of an obstacle course used for navigation testing. Bottom – The prototype wearable system for navigation.

Object detection system testing used a prosthetic vision simulator, described previously.[7] Briefly, a webcam was mounted on a head-worn display (Arrington Research, Inc.). The video stream from the webcam was modified using custom software and rendered as a simulation of prosthetic vision with a controllable format in terms of number of pixels active, pixel location, and pixel size. Each pixel represented an active electrode of a retinal prosthesis. In addition to the grid of active pixels, cuing pixels were placed at 8 locations: up, down, left, right, up-left, up-right, down-left, and down-right. The video stream was simultaneously processed by a saliency algorithm that detected the most important object in the scene.[8] Briefly, the saliency algorithm detects salient parts of an image, based on color, intensity, and orientation versus the background. These algorithms are modeled on primate vision. Only areas outside the view of the user, but within the view of the camera were processed for important objects. The cueing pixels directed the user to look in the direction of the most important object.

Subjects were asked to locate up to three objects on an otherwise empty tabletop. Subjects performed the tasks with and without cues. 7 normally sighted subjects were enrolled for this study. The subjects were provided with simulated prosthetic vision in the central diagonal 10 degrees of the HMD. Prior to testing, the subjects were allowed to freely view object on the table to gain experience with the system. Subjects wore a shroud over the HMD to block their peripheral vision and were seated at a desk. 1, 2 or 3 objects were placed on the desk at a time and the subjects were asked to find them using the simulated vision. The only objects on the desk were the objects of interest and one central reference object. Subjects first performed the task with no cues and then with cues. Each subject performed 6 trials of the 1, 2 and 3 object cases, 3 trials with cues and 3 trials without cues. The data thus consisted of 18 trials from each subject and a total of 126 trials from all subjects for both cases. Data was recorded for total head movements in the horizontal and vertical directions, the number of errors and the time taken to finish the task.

Figure 2 shows a typical arrangement for the objects for this experiment. The round tape roll in the center of the desk acted as a starting point and reference for the subjects. The subjects were instructed to fixate on the center object, find all the objects on the table, then return to the center. In the beginning of each trial, to prevent subjects from finding objects while trying to find the center starting point, the HMD was disabled. The subjects were guided verbally until the camera was fixated on the center starting point, and then the HMD was enabled. For the no cueing trials, the subjects scanned the desk to find the object(s). For the cueing trials, the subjects looked at the central reference and were provided with directional cues from the algorithm. For the cueing trials having more than one object placed on the table, the subjects made head movements towards each object and came back to the central reference to wait for the next directional cue to find the other object(s).

All human subjects testing was approved the University of Southern California Institutional Review Board.

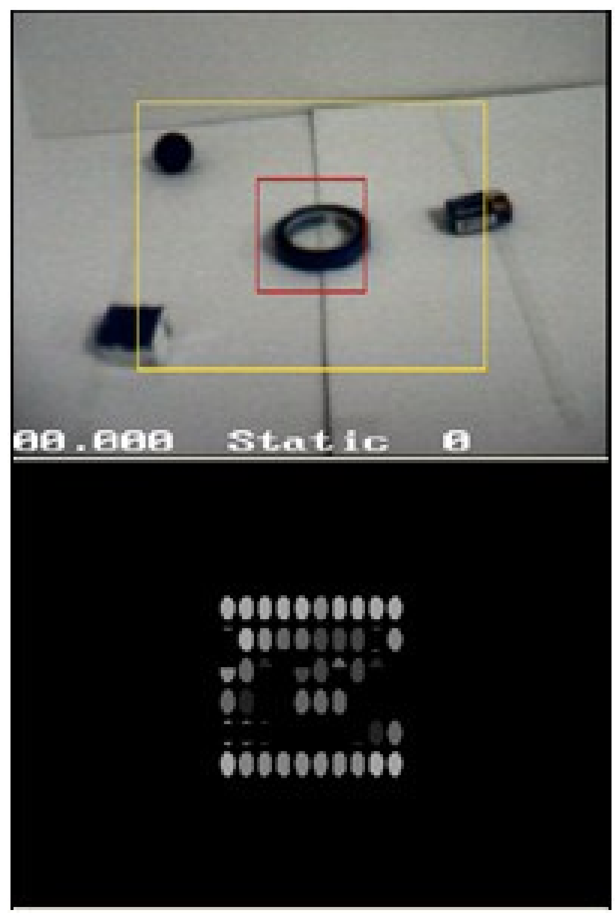

Fig. 2 – Top – yellow box indicates the area processed by the saliency algorithm, red box shows the area that is pixelized. Bottom – simulated retinal prosthesis pixels

III. Results

Group 3 (cane and device) subjects had no collision in any trial. Group 1 had 17 collisions and group 2 totaled 7 collisions (Figure 3) Thus, fewer collisions occurred when using the device. The difference was maintained when groups 1 and 2 were swapped. Video recordings were used to create 'heatmaps' to visualize trajectories adopted by the subjects. These heatmaps demonstrated that white cane users contacted sides of obstacles and got stuck behind longer objects due to the limited sensing range of the cane. However, the average Percentage Preferred Walking Speed (PPWS) measured for the 3 groups across all trials showed that group 1 was fastest followed by group 3; group 2 was slowest. Performance with the device-only was slow due to limitations of the prototype system and because subjects overcompensated when cues were received. In some cases, these delays led to disorientation, which was overcome by group 3 subjects who also used the cane to maintain a sense of direction.

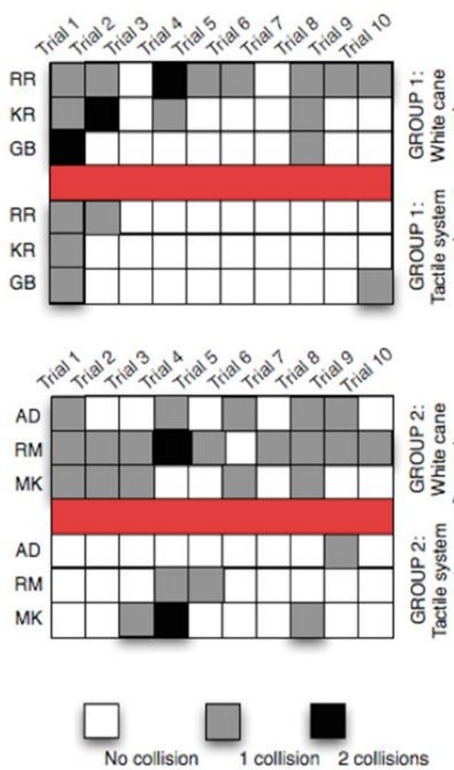

Figure 3 – Number of collisions for groups 1 and 2. Group 3 had no collisions

Figures 4 shows the time in seconds for each trial averaged over the 7 subjects for the 1, 2 and 3 object cases respectively for both the no cueing and cueing trials. The time taken to finish the trials with cues was significantly less than the time taken to complete the NC trials (paired t-test, $p<0.05$).

IV. Conclusions

The experiments demonstrate that computer vision algorithms for navigation and object detection may benefit the visually impaired. Such algorithms could be used with a retinal prosthesis or as a wearable system. In the latter case, acceptable human interface hardware will be needed to communicate information via either sound or tactile cues.

Acknowledgment

We acknowledge Dr. Armand Tanguay Jr. and Ben McIntosh for developing the retinal prosthesis simulator.

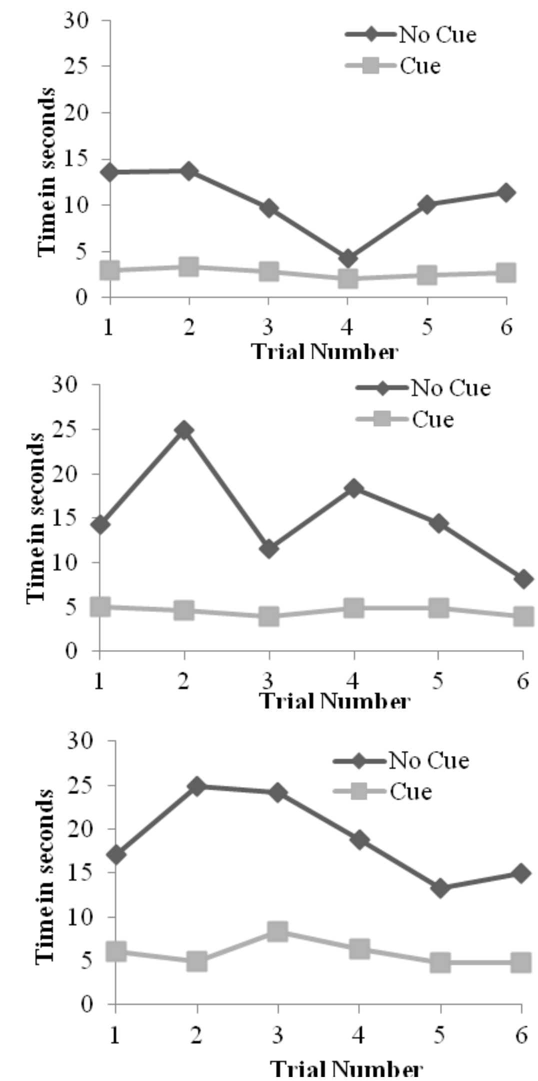

Figure 4: Time in seconds for the 1, 2 and 3 object cases in (a), (b) and (c) respectively for both no cueing and cueing groups

REFERENCES

[1] Weiland JD, Liu W, Humayun MS. Retinal Prosthesis. Annu Rev Biomed Eng. 2005;7:361-401.

[2] Ptito M, et al. Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind, Brain (2005) 128 (3): 606-614.

[3] Amedi A, et al. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience* **10**, 687 - 689 (2007)

[4] Humayun MS, Dorn JD, da Cruz L. et. al. Interim results from the international trial of Second Sight's visual prosthesis. *Ophthalmology.* 2012 in press.

[5] Zrenner E, Bartz-Schmidt KU, Benav H, Besch D, Bruckmann A, Gabel VP, Gekeler F, Greppmaier U, Harscher A, Kibbel S, Koch J, Kusnyerik A, Peters T, Stingl K, Sachs H, Stett A, Szurman P, Wilhelm B, and Wilke R. Subretinal electronic chips allow blind patients to read letters and combine them to words. Proc Biol Sci. 2011 May;278(1711):1489-1497.

[6] Pradeep V, Medioni G, Weiland JD. A Wearable System for the Visually Impaired. International IEEE Engineering in Medicine and Biology (EMBS) Conference 2010

[7] Parikh N, McIntosh BP, Tanguay Jr AR, Humayun MS, Weiland JD. Biomimetic image processing for retinal prosthesis: Peripheral saliency cues. IEEE Eng Med Biol Soc. 2009 pages 4569-4572.

[8] Parikh N, Itti L, Weiland JD. Saliency-based image processing for retinal prostheses. J Neural Eng. 2010 Feb;7(1):16006. Epub 2010 Jan 14.PMID: 20075505