GOING BEYOND VISION TO IMPROVE BIONIC VISION

Wai Ho Li*, Titus Jia Jie Tang, Wen Lik Dennis Lui
Monash Vision Group, Melbourne, Australia
*waiholi@gmail.com

ABSTRACT
Currently, most implanted visual prosthetic systems generate vision by translating sensor data from a headworn camera into electrical stimulation of the human vision system. Unfortunately, the resulting bionic vision has low spatial resolution and limited dynamic range. This dramatically reduces the usefulness of bionic vision in many real world scenarios. Historically, this problem is treated as immutable pathology. Recently, image processing has been proposed as a potential remedy to improve the useability of bionic vision. We explore another alternative: Combining multiple sensing modalities and robotic sensing algorithms. This paper gives a top level summary of ongoing research exploring this alternative.

Index Terms— Visual prosthesis, Computer vision, Robot sensing systems, Human factors

1. INTRODUCTION

Monash Vision Group is a research center with the aim to develop a cortical visual prosthesis, which produces bionic vision by electrically stimulating a patient’s primary visual cortex (V1). Electrical stimulation of V1 elicits spots of light, phosphenes, at predictable locations in a patient’s visual field. Our device will stimulate V1 using multiple wireless intracortical electrode arrays, resulting in a 2D grid of phosphenes similar to a low resolution binary image with around 500 individual pixels. An overview of our cortical visual prosthesis is available from a paper in the same ICIP special session [1].

Figure 1 provides a top level system diagram of a typical implanted visual prosthesis. Currently, the two main regions of the human vision system considered for implantation are the retina and V1. In order to evaluate visual prostheses, Simulated Prosthetic Vision (SPV) trials are conducted by showing sighted volunteers bionic vision visualisations via a Head Mounted Display (HMD). Further details of our group’s SPV trial procedures are available from a paper in the same ICIP special session [2].

Figure 2 shows a simulation of the bionic vision that maybe possible with our device. The combination of simple image processing using area-based down sampling and binary thresholding together with the low resolution and dynamic range of bionic vision truncates much of the salient content from the scene. For example, the contours of the objects on the table have been lost. The usual problems caused by shadows, texture and specular reflections when applying image processing algorithms is amplified by the resolution and dynamic range constraints of bionic vision.

Advanced image processing has been previously suggested as a way to improve the useability of bionic vision [3]. For example, a scene’s contrast can be artificially enhanced using histogram-based normalisation. This paper investigates alternative approaches that improves bionic vision by going beyond vision-only approaches; using non-visual sensors, multiple sensing modalities and robotic sensing algorithms.

*This work was funded by the Australian Research Council Special Research Initiative in Bionic Vision and Sciences (SRI 1000006)
2. TRANSFORMATIVE REALITY

In 2011, we proposed a novel concept called Transformative Reality (TR) to improve bionic vision. This work has led to a US provisional patent application at the end of 2011 followed by a PCT filling in 2012. This section provides a summary of TR. The interested reader can obtain more details from our previous papers [4, 5].

Transformative Reality (TR) can be viewed as a specialized subset of Steve Mann’s mediate reality (MR) conceptual framework [6]. Essentially, we treat the implanted visual prosthesis as a technology that mediates the real world to a user through low resolution bionic vision. Recall the problematic scene in Figure 2. Instead of restricting ourselves to a strict, direct visual representation of a scene, TR asks the following question: What combination of sensors, processing and rendering gives the most useability? Instead of using a colour camera, what if we use a range camera and accelerometer? What if we represent the world symbolically, such as using avatars to represent human faces (Figure 5) to get around the limited resolution in the bionic vision output?

The following TR modes were designed with the goal of improving task performance for common daily activities; activities that a focus group of vision impaired users nominated as challenging to perform using traditional assistive technologies such as a cane or guide dog. These tasks include object detection and recognition, navigation of cluttered indoor spaces and interacting with people.

Figure 3 shows a TR mode we call Structural Edges, which highlights the 3D discontinuities and creases of a scene captured using a range camera. This is done by looking for non-planar local regions in the depth image. SPV user trials suggest that this TR mode provides noticeable improvements for object detection and recognition compared to the simple approach shown earlier in Figure 2.

![Fig. 3. Structural Edges TR mode. The depth data on the left is taken from the scene on the LHS of Figure 2.](image)

Figure 4 shows the Empty Ground TR mode, which was designed to help patients navigate cluttered indoor areas. This TR mode renders patches of Empty Ground as lit phosphenes. The algorithm works by detecting the ground plane using RANSAC plane fitting, which can be done at frame rate by limited the search space to planes with normals opposing the direction of gravity. Gravity is estimated using an accelerometer. SPV user trials suggest that this TR mode allows safe navigation of cluttered indoor areas.

![Fig. 4. Empty Ground TR mode.](image)

Figure 5 shows the People TR mode, an early proof of concept, which highlights people in a scene. This mode makes use of a colour camera and a range camera. Firstly, a face is detected in the colour camera image (Viola-Jones). The 3D range of the face is then found using the range camera. This is followed by a search for a contiguous blob below the face, which is likely to be a person’s body. The face is rendered symbolically using an avatar to effectively represent a person’s face in low resolution. The body is rendered as a contiguous blob of phosphenes. Currently, this mode only shows people facing the patient as it relies on frontal faces.

![Fig. 5. People TR mode.](image)

All three TR modes shown above runs at frame rate on consumer PC hardware. A single RGB-D sensor, the Microsoft Kinect, provides the required sensor data to the TR algorithms. To fully appreciate the useability of each TR mode in dynamic real world scenarios, we highly recommend viewing the following video: [http://youtu.be/J30uYYkDApY](http://youtu.be/J30uYYkDApY)

3. DETECTING PLANES AND STAIRCASES

Planar regions are prevalent in many indoor scenes and represent useful navigational landmarks (doors, windows, stairs, tables, chairs). Planar regions can also act as support surfaces for other objects. Here, we summarize research investigating the use of planes for staircase detection. Details of our algorithm are available from [7].

We begin by defining a staircase as a series of ascending or descending planar regions anchored to a ground plane. Our definition requires an observation of the ground plane in order...
to function; a limitation that we are planning to remove in the near future. Note that similar to the Empty Ground TR mode, our staircase detector uses an accelerometer to achieve real time performance by limiting planar search space based on the direction of gravity.

Our algorithm starts by searching for a ground plane based on an estimate of the height of the headworn sensor when worn by a walking patient. After finding the ground plane, the algorithm then searches for the first step above or below the ground plane. This search is repeated in a consistently upward or downward manner depending on the staircase’s inclination as determined by the very first step found by our algorithm. This process is shown in Figure 6.

![Fig. 6. Detecting staircases as a series of planes. The blue pixels are the inlier depth pixels of the detected ground plane. As the depth and colour cameras have offset optical centers, there is a slight shift in the way the pixels are overlaid. The results of staircase detection is shown on the right. Note that the underlying step model is an infinite plane in 3D space.](image)

A positive side effect of using an active range camera, such as the Microsoft Kinect, is that these sensors measure depth by illuminating the scene. This means that our system is able to detect staircases in dark environments, such as the scene in Figure 7.

![Fig. 7. Staircase detector working in the dark.](image)

In order to test the algorithm in a more realistic setting, we built a custom head mounted system to collect data from a patient’s point of view. This prototype system is shown in Figure 10. Our algorithm was evaluated on a data set collected using our prototype system. The data set consists of 121 datum (Colour image, Depth image and Accelerometer readings).

Overall, our system achieved a false positive rate of 1.02% and false negative rate of 5.07%. The system performed well for both ascending and descending staircases. Our algorithm is also able to report the inclination and the number of steps of a staircase in real time to the user. A detailed breakdown of system performance is available in [7]. A video of the staircase detector in action is available here: http://youtu.be/zY6JitogSuk

4. EGOMOTION ESTIMATION

Egomotion estimation is the process of keeping track of the motion of a moving sensor in a previously unmodelled environments. This section describes an egomotion estimator that provides a 6 Degrees of Freedom (DoF) estimate of a range camera’s location in real time. Details of the algorithm and its performance, including a novel inverse depth formulation of ICP, are available from [8].

Egomotion estimation allows us to track the motion of a headworn sensor, which may provide useful information such as whether a patient is revisiting a location or looking at the same part of a scene. For example, Figure 8 shows egomotion results where a sensor was moved by hand around a cluttered table in a circular path. A video demonstration of our egomotion estimation method is available here: http://youtu.be/4ae_kHeygIw

![Fig. 8. Egomotion estimation of a handheld sensor moved around a tabletop scene. The airplane represents the 6-DoF orientation of the handheld sensor.](image)

Figure 9 shows a 3D reconstruction of a scene built by moving the range camera using a turntable, which simulates a slow head turn. This kind of reconstruction allows small local maps of the world to be built by a visual prosthesis system, potentially enabling the conveyance of useful information that is not in the sensor’s current field of view. An important consideration given the wide angle field of view of human vision in comparison to most range and vision sensors.
5. DATA COLLECTION AND SIMULATED PROSTHETIC VISION

Several head mounted systems were prototyped to aid in the collection of sensor data and to enable Simulated Prosthetic Vision trials. The system in Figure 10 was used to collect data to evaluate the plane and staircase detectors. A Kinect is rigidly mounted on the front of a safety helmet, which allows the system to be worn comfortably over long periods of time. An optional head mounted display (VR goggles) can be worn to visualise input data and algorithm outputs.

![Kinect and Display System](image)

**Fig. 10.** Head mounted system for data set collection

We have also developed a more lightweight system as shown in Figure 11. This system is used primarily for agile TR algorithm prototyping as it enables mobile Simulated Prosthetic Vision (SPV) trials. This system replaces an older and more cumbersome head mounted display system, which can be seen in the following SPV trial video: [http://youtu.be/iK5ddJqNuxY](http://youtu.be/iK5ddJqNuxY)

![Lightweight Head Mounted System](image)

**Fig. 11.** Head mounted system used for TR development

6. CONCLUSION

Low cost range cameras are now commodity items. Such sensors are becoming so small and lightweight that they are being embedded into mobile devices (PrimeSense Capri, PMD Tec PhotonIC). As we have shown in this survey paper, range cameras add value to visual prostheses above that provided by a colour camera. Fusing data from multiple sensors and using robotic sensing algorithms have the potential to improve the usefulness of implanted visual prostheses despite the resolution and dynamic range constraints of bionic vision.

Future work includes Simulated Prosthetic Vision trials to further evaluate our algorithms and extending the egomotion research to a Simultaneous Localisation and Mapping (SLAM) framework capable of building indoor maps.

7. REFERENCES


