35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Enhancing scene structure in prosthetic vision using iso-disparity contour perturbance maps

David Feng and Chris McCarthy

Abstract— **We present a novel approach for enhancing structurally significant features in a scene to facilitate safe mobility with prosthetic vision. Previous approaches rely on visually salient features (*e.g.*, intensity gradients, size, texture), or surface fitting (*e.g.*, ground plane extraction), to determine and convey regions of structural change in the scene. Such approaches can be costly to compute, and/or are not guaranteed to detect all features relevant to the needs of safe mobility (*e.g.*, small, low-contrast trip hazards). Assuming a dense disparity image, we propose a novel feature using iso-disparity contours. Regions of significant structural change are detected via a cost function based on local comparisons of iso-disparity contour orientations. Through this, structurally interesting features such as surface boundaries and general clutter are extracted and emphasised in the output visual representation. Our approach is real-time, and requires no surface fitting. Experimental results quantitatively and qualitatively validate our approach.**

I. INTRODUCTION

Enabling safe and efficient mobility is a primary aim of current and near-term visual prostheses. In particular, retinal prostheses have seen significant advances in recent years. Retinal prostheses achieve stimulation via an array of electrodes which is implanted close to the retina. Electrical stimulation aims to elicit a neural response in the retinal ganglion cells, leading to higher levels of response in the visual cortex. The percept elicited by this process is known as a *phosphene*: described as a bright 'star-like' spot of light [1]. Psychophysical studies show that the shape and brightness of phosphenes can be varied by modulating stimulation parameters, allowing visual representations of the scene to be rendered. However, current devices are significantly restricted in the resolution and dynamic range they provide, motivating researchers to consider ways to efficiently encode visual information about the scene.

Most visual prostheses acquire scene information via an externally worn camera (exceptions include [2], where eye-resident photodiodes are used). This allows vision processing to extract important information present in the high resolution images and encode it in efficient visual representations of the scene appropriate for the implant.

In the case of orientation and mobility with prosthetic vision, previous work has primarily focussed on mobility using down-sampled intensity images. Studies such as [3], [4], [5], [6], [7] have demonstrated basic way-finding and orientation using intensity alone with relatively few phosphenes. These studies, however, assume (or construct) high-contrast environments to assist navigation. Visual saliency has also been explored, both for cueing obstacles in the visual representation [8], or using the output saliency map as the visual representation [9]. The use of intensity features such as edges, gradients, and texture have a strong biological basis, however, they are unlikely to be sufficient for safe mobility away from high contrast conditions with current and near-term implants. In particular, small low-contrast obstructions on the ground surface are likely to be missed.

Previous work has shown that artificially enhancing the contrast between obstacles and the ground plane can significantly improve the perception of small ground-based obstacles [10]. In [11], this is achieved by extracting a ground-plane model in stereo disparity data in order to augment a depth-based representation by darkening the ground, and scaling up all non-ground phosphenes (referred to as *Augmented Depth*). However, surface fitting is error prone, and does not easily scale to complex, cluttered scenes. Determination of the ground plane can also be ambiguous, often requiring dominant surface assumptions which do not always hold when the camera is head-mounted, and scanning the scene.

In this paper we propose a novel method for enhancing structurally important features in the scene. Unlike previous approaches, we achieve this without computing pixel-wise surface normals, estimating surface models or use of appearance-based features in colour/intensity images. Rather, we exploit the appearance of iso-disparity contours, *i.e.*, lines representing level sets of disparity, in disparity images to statistically determine regions of structural significance in the scene (*i.e.*, surface boundaries and general clutter). We have previously reported the use of iso-disparity contours for planar surface fitting [12]. Here, we do not explicitly model surfaces, but instead treat iso-disparity contour orientations as an observable feature in disparity space, from which smooth and non-smooth regions may be inferred. Results show our method accurately and robustly highlights all obstructions in the scene, as well as major surface boundaries. Qualitative examination of the resulting visual representation in simulated prosthetic vision demonstrates the potential of our approach to support safe mobility with current and near-term visual prostheses.

All authors are with NICTA Canberra Research Laboratory, Canberra ACT, Australia and College of Engineering and Computer Science, Australian National University, Canberra ACT, Australia.

The authors acknowledge the contributions of Paulette Lieby, Adele Scott, Ashley Stacey, Jeremy Oorloff and Nathan Gough, authors of the simulated prosthetic vision software used in this paper. This research was supported by the Australian Research Council (ARC) through its Special Research Initiative (SRI) in Bionic Vision Science and Technology grant to Bionic Vision Australia (BVA). NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

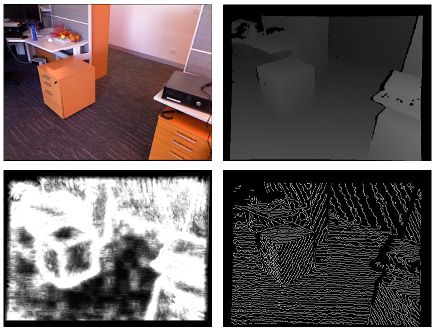

Fig. 1. Clock-wise from top left: rgb image of scene, input disparity image, iso-disparity contour image, perturbance map.

II. Approach

Input is a dense, discretised disparity image D, obtained by inverting and scaling the depth map obtained from an RGB-D sensor. The described process assigns each pixel in D a 'perturbance' score, which reflects the structural significance of a given region such that pixels corresponding to clutter or surface boundaries are expected to have a higher perturbance score than pixels on smooth surfaces. The perturbance score is calculated by locally comparing the orientation of iso-disparity contours in local regions. By definition, smooth surfaces will exhibit highly uniform iso-disparity contours. In contrast, clutter and surface boundaries will exhibit relatively non-uniform and/or discontinuous iso-disparity contours. Based on this observation, we detect such features in four steps, outlined below:

A. Extraction and Multi-scale histogramming of iso-disparity contours

Canny edge detection [13] is applied to D to produce a binary image of iso-disparity contours, determined from the boundary between discrete disparity levels (see Figure 1). These iso-disparity contours are then divided into linear piecewise segments, in order to estimate the local orientation of each contour point. This is achieved by iteratively forming straight line segments on contour points until an error of 4 pixels is exceeded, at which point the segment is stored and the process repeated.

A multi-scale sliding window is passed over the iso-disparity image to determine the local distribution of iso-disparity orientations at each position. Orientations within each window are counted into a number of discrete histogram bins B. In the experiments we set $B = 9$. For a window with side length s at position (u, v), the resulting histogram is denoted as $H_{s,u,v} : [1 \,..\, B] \rightarrow \mathbb{R}$.

B. Window perturbance calculation

The dissimilarity between two windows in the same scale is computed by comparing their histograms. We define a smoothed cost metric between two windows of scale s positioned at (u, v) and (u', v') as

$$\mathcal{C}_s(u,v,u',v') = \log\left(1 + \frac{\sum_{b=1}^{B} \left|H_{s,u,v}(b) - H_{s,u',v'}(b)\right|}{\sum_{b=1}^{B} \max(H_{s,u,v}(b), H_{s,u',v'}(b))}\right)$$

This cost reflects the non-overlapping portion of the histograms as a ratio to total size. A nonlinear function is used to balance and reduce the effect of large costs.

The perturbance of a given window is taken as the minimum cost between the window and neighbouring windows in the same scale. Thus, given a scale s, window position (u, v), neighbourhood radius r, and neighbour step size ns, the scale perturbance of the window is given by

$$\mathcal{P}_s(u,v,r,ns) = \min_{-r \le i \le r, -r \le j \le r} \mathcal{C}_s(u,v,u+i\cdot ns, v+j\cdot ns)$$

Taking the minimum effectively compares the window to its most similar neighbour. The dissimilarity between the two is proportional to the likelihood that the pixel represents clutter. In the experiments, we set $r = 10$ and $ns = 15$.

C. Multi-scale perturbance calculation

The multi-scale perturbance of a given point is computed by first finding the window perturbance of the point for a number of different scales, and then merging the results via a weighted average by window occupancy. Thus, defining S as the set of window sizes and $C(u, v, s)$ as the iso-disparity contour pixel count of the window of size s positioned at (u, v), the perturbance score for a pixel is given by

$$\mathcal{P}(u,v,r,ns,S) = \sum_{s \in S} \frac{C(u,v,s)}{s \cdot s} \mathcal{P}_s(u,v,r,ns)$$

We set S as $\{20, 30, 40\}$ in the experiments.

D. Contour-disparity ratio and gradient magnitude adjustment

We perform two post-processing steps on the perturbance image using D: lowering the perturbance of fronto-parallel surfaces, and increasing the perturbance of depth-discontinuity edges.

Fronto parallel surfaces pose a challenge since their iso-disparity contours are relatively sparse. We detect such surfaces explicitly by computing the local ratio of iso-disparity pixels to the total number of valid disparity values. If the ratio is near-zero, then we assume the region represents a near-frontal surface.

Gradient magnitude thresholding of D was performed to explicitly identify depth discontinuity edges in order to increase surface boundary recall. Any pixel in the normalised gradient magnitude image of D with a value above 0.0005 was assumed to represent part of a depth discontinuity edge.

III. Experimental Results

A. Quantitative comparison: surface boundary recall

To validate the appropriateness of our approach, we provide two measures:

- surface boundary recall rate (SBRR): the proportion of correctly labelled pixels along surface boundaries; and,
- ground plane mislabel rate (GPMR): the proportion of ground plane pixels incorrectly labeled as surface boundaries/clutter.

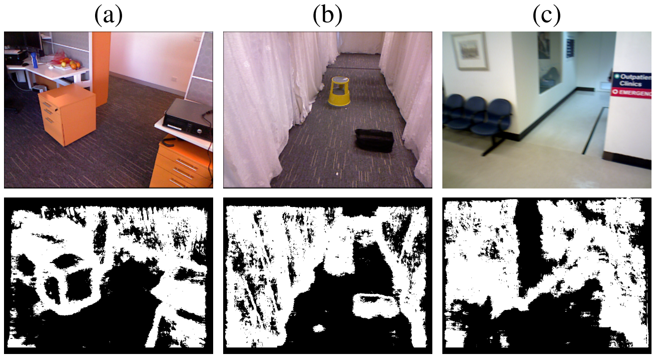

Fig. 2. Images used for quantitative results in Table I, and their corresponding t=0.4 binary thresholded perturbance maps.

	Perturbance t=0.4		Perturbance t=0.7		Plane Fitting	
Image Id	SBRR	GPMR	SBRR	GPMR	SBRR	GPMR
(a)	0.98	0.18	0.62	0.04	0.74	0.00
(b)	0.94	0.21	0.54	0.05	0.52	0.05
(c)	0.90	0.25	0.52	0.07	0.71	0.03

TABLE I

QUANTITATIVE RESULTS SHOWING SURFACE BOUNDARY RECALL RATE (SBRR), AND GROUND PLANE MISLABEL RATES (GPMR) FOR THE PERTURBANCE MAP (THRESHOLDS 0.4 AND 0.7), AND PLANE FITTING.

Ground truth was obtained via hand-labelling of pixels along all surface boundaries, and the ground plane. These metrics were calculated using a thresholded binary segmentation of the normalised perturbance image.

Table I shows results for a set of test images shown in Figure 2. Here we report SBRR and GPMR results using the proposed perturbance map with thresholds $t = 0.4$ and $t = 0.7$. For comparison, we also include results obtained from the plane-fitting technique described in [12]. We include this to validate how the proposed method compares in distinguishing navigable vs non-navigable space in the scene.

Most notably, the 0.4 perturbance map achieves an SBRR above 90% for all images (*i.e.*, $98\%, 94\%$, and 90%), with the best result achieved for Figure 2(a). GPMR results for the perturbance map are less impressive for t=0.4 (*i.e.*, $18\%, 21\%$, and 25%), but improve significantly for t=0.7 (*i.e.*, $4\%, 5\%$, and 7%), indicating a clear trade-off between recall rate and mislabelling. Visual inspection of the 0.4 perturbance segmentation shows that in all images, mislabelling is primarily due to the thickness of boundary segmentations. Away from the ground surface boundaries, mislabelling is rare. The comparatively lower SBRR results for plane fitting are unsurprising given the method makes no explicit attempt to detect boundaries. Thus, ground plane labels can easily bleed across boundaries.

B. Qualitative assessment

The first two columns of Figure 3 show sample images and the resulting perturbance map obtained using the proposed method. It can be seen that the perturbance map provides clear delineation between clutter in the scene and the dominant smooth surfaces. In particular, small ground obstacles such as those shown in rows 1 and 4 are given high perturbance scores relative to the ground plane. Non-ground smooth surfaces such as walls (Row 3) and table tops (Row 5) are de-emphasised relative to other clutter in the scene.

Column 3 shows a our proposed perturbance-based visual representation using simulated prosthetic vision (SPV)[1] Phosphene levels are determined from a direct sampling of the smoothed and normalised perturbance map, thus conveying the extent of clutter and non-smoothness in each each phosphene's visual field. For comparison, Column 4 shows an SPV rendering using a standard intensity-based visual representation (*i.e.*, down-sampling of the original intensity image). Column 5 shows the plane-fitting-based *Augmented Depth* visual representation [11], in which a depth-based visual representation is augmented to increase contrast between ground and non-ground phosphenes using ground plane segmentation. While providing similar artificial enhancement of ground obstacles to our proposed method, we expect the perturbance-based approach to provide a more general perception of structure in the scene. This increased emphasis of scene structure is particularly evident in rows 2, 3 and 5 where obstacles are similarly emphasised, but more distinguishing detail is present in the perturbance-based visual representation. As expected, low-contrast objects are generally not visible in the intensity-based representation.

IV. Discussion

The above results demonstrate the effectiveness of the proposed perturbance map for detecting and emphasising structurally significant regions in the scene. While plane-fitting methods can generally be expected to achieve greater ground pixel labelling than our approach, quantitative results above demonstrate a significantly better recall rate for surface boundaries using the perturbance map. This is arguably the more relevant performance indicator for safe mobility with prosthetic vision, ensuring all boundaries in the scene are preserved, and no potential trip hazards are missed. It is also important to note that the perturbance-based approach makes no planar surface assumptions; it simply characterises *smoothness*. The perturbance map also provides a richer description of the scene, characterising all regions of clutter, as well as object shape. A possible extension of this work is to combine the perturbance map and plane fitting under a globally optimised segmentation frame-work.

V. Conclusion

We have proposed a novel feature for detecting and characterising regions of structural interest in the scene to support mobility with a visual prosthesis. Our approach avoids assumptions of high contrast environments, and removes the need to explicitly reconstruct the scene via surface fitting. Our results demonstrate how the proposed perturbance map may be utilised to emphasise all surface boundaries and clutter in the scene, robustly and efficiently. More generally, the proposed approach demonstrates how a more qualitative

[1] We show these results using a simulation based on Bionic Vision Australia's suprachoroidal 98 electrode array [14].

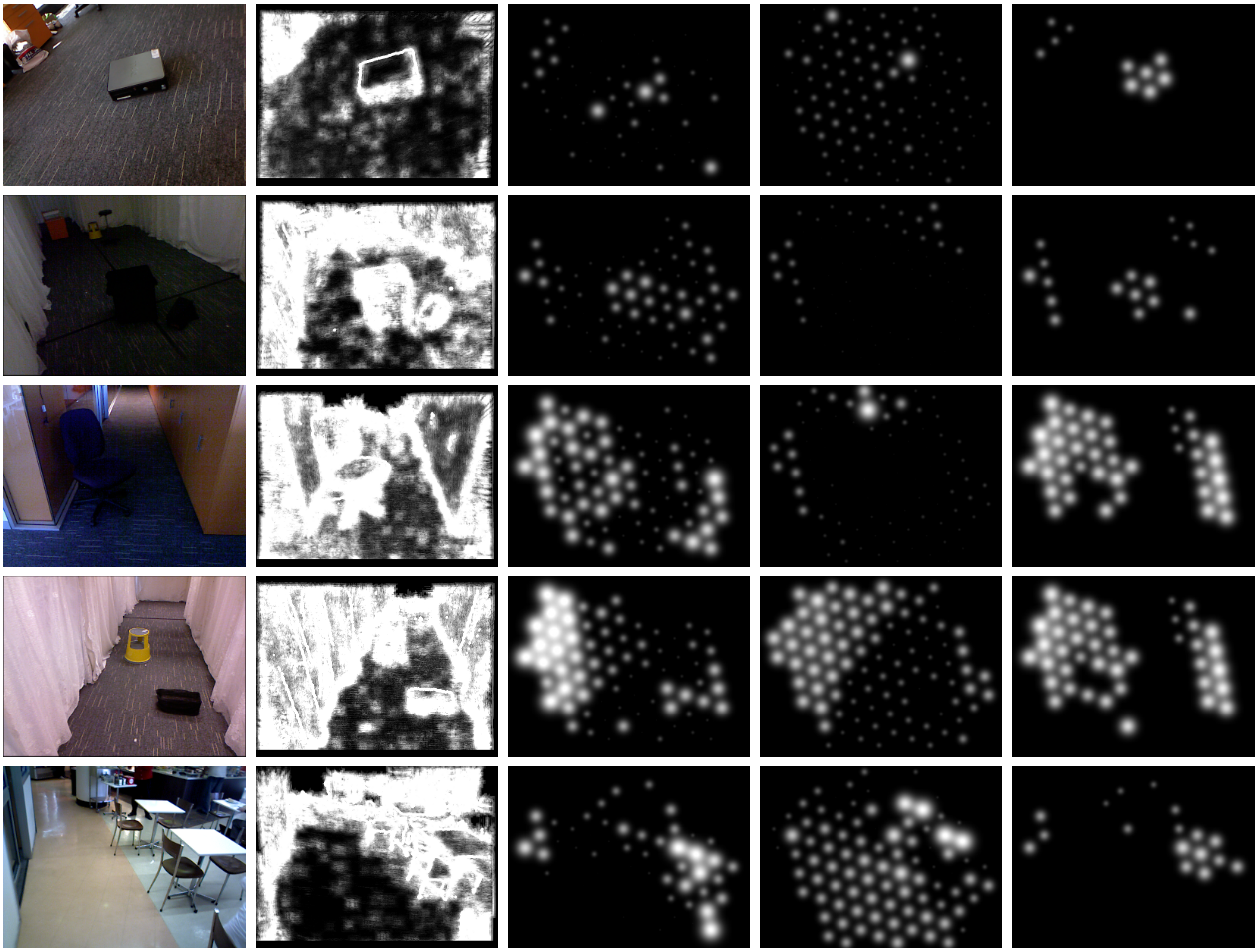

Fig. 3. Qualitative results (from left to right): RGB image, perturbance map, perturbance-based SPV, intensity-based SPV, and Augmented Depth[11] SPV

analysis of scene structure using depth data may provide advantages for supporting mobility with near-term prosthetic vision devices. Human mobility trialling must be conducted to more accurately assess this.

References

[1] G. Brindley and W. Lewin, "The sensations produced by electrical stimulation of the visual cortex," *The Journal of Physiology*, vol. 196, no. 2, p. 479, 1968.

[2] E. Zrenner, K. U. Bartz-Schmidt, H. Benav, D. Besch, A. Bruckmann, V.-P. Gabel, F. Gekeler, U. Greppmaier, A. Harscher, S. Kibbel, J. Koch, A. Kusnyerik, T. Peters, K. Stingl, H. Sachs, A. Stett, P. Szurman, B. Wilhelm, and R. Wilke, "Subretinal electronic chips allow blind patients to read letters and combine them to words," *Proceedings of the Royal Society B: Biological Sciences*, vol. 278, no. 1711, pp. 1489–1497, 2011.

[3] K. Cha, K. W. Horch, and R. A. Normann, "Mobility performance with a pixelized vision system," *Vision Research*, vol. 32, no. 7, pp. 1367 – 1372, 1992.

[4] G. Dagnelie, P. Keane, V. Narla, L. Yang, J. Weiland, and M. Humayun, "Real and virtual mobility performance in simulated prosthetic vision," *Journal of Neural Engineering*, vol. 4, pp. S92–S101, 2007.

[5] D. J. A. and A. J. Maeder, "Mobility enhancement and assessment for a visual prosthesis," in *SPIE Medical Imaging 2004: Physiology, Function, and Structure from Medical Images*. International Society for Optical Engineering, 2004.

[6] J. Dowling, W. Boles, and A. Maeder, "Mobility assessment using simulated artificial human vision," in *Proceedings of the 2005 Workshop on Computer Vision Applications for the Visually Impaired (CVAVI)*, june 2005.

[7] M. S. Humayun, L. da Cruz, G. Dagnelie, S. Mohand-Said, P. Stanga, R. N. Agrawal, and R. J. Greenberg, "Interim performance results from the second sight Argus II retinal prosthesis study," in *Proceedings of the Association for Research in Vision and Ophthalmology (ARVO)*, 2010.

[8] M. S. H. N. Parikh and J. D. Weiland, "Mobility experiments with simulated vision and peripheral cues," in *Proceedings of the Association for Research in Vision and Ophthalmology (ARVO)*, 2010.

[9] J. Boyle, A. Maeder, and W. Boles, "Region-of-interest processing for electronic visual prostheses," *Journal of Electronic Imaging*, vol. 17, no. 1, pp. 013 002–013 002, 2008.

[10] C. McCarthy, P. Lieby, J. G. Walker, A. F. Scott, V. Botea, and N. Barnes, "Low contrast trip hazard avoidance with simulated prosthetic vision," in *Proceedings of the Association for Research in Vision and Ophthalmology (ARVO)*, 2012.

[11] C. McCarthy, N. Barnes, and P. Lieby, "Ground surface segmentation for navigation with a low resolution visual prosthesis," in *Proceedings of IEEE EMBC 2011*. IEEE, 2011, pp. 4457–4460.

[12] C. McCarthy and N. Barnes, "Surface extraction from iso-disparity contours," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2010.

[13] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–714, 1986.

[14] Y. Wong, S. Chen, J. Seo, J. Morley, N. Lovell, and G. Suaning, "Focal activation of the feline retina via a suprachoroidal electrode array," *Vision Research*, vol. 49, no. 8, pp. 825 – 833, 2009.