33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Ground surface segmentation for navigation with a low resolution visual prosthesis

Chris McCarthy, Nick Barnes and Paulette Lieby

Abstract— **We propose the use of ground surface segmentation to enhance the perception of obstacles in low to medium resolution prosthetic visual representations. We apply a recently proposed algorithm for segmenting traversable space in stereo disparity data, and show how such a scheme may be utilised to enhance the distinction between the ground surface and obstructions (in particular, small trip hazards). Qualitative comparisons with intensity and straight depth-based representations highlight advantages for the visualisation of obstacles, offering potential gains for visual navigation with low resolution and low dynamic range visual prostheses.**

I. INTRODUCTION

Bionic Vision Australia seeks to develop a retinal implant capable of supporting visual ambulatory navigation. The envisaged resolution and dynamic range of current and near-future visual prostheses suggests this will require more than down-sampled images of scene luminance alone [1]. More efficient representations of environmental structure making maximal use of stimulation points are likely to be required. To support this, salient features in the scene will need to be extracted and incorporated into the stimulation strategy. Current trends (based on weight/cost/bandwidth trade-offs) suggest digital image data will be the primary source for obtaining such information [2], [3].

The perceptual response elicited by a single electrode is referred to as a *phosphene*, and is generally described as resembling a bright 'star-like' spot of light [4]. Psychophysical studies have verified that the size and luminance of each phosphene can be modified by varying the current and frequency of the associated electrical stimulus [5]. An accepted methodology for studying the functional capacity of prosthetic vision is via simulated phosphene vision (SPV, see [5], [1] for a review). In this, digital images are *phosphenised* to simulate the visual conditions of prosthetic vision for sighted participants.

Obstacle avoidance is critical to ambulatory navigation. While studies such as [6], [7], [8], [9], [10], [11] provide empirical support for basic obstacle avoidance with SPV using scene luminance (referred to as *intensity-based*), the experimental conditions generally provide high contrast change across surface boundaries to support this. Recent work has reported the use of surface depth as an alternative scene representation [12], [13]. This is achieved by computing stereo disparities between two space-separated parallel cameras, and modulating phosphene size and intensity in accordance with the measured proximity of surfaces along the phosphene's mapped viewing direction (*i.e.*, bright means close, darker means further away). Results demonstrate the ability of sighted participants to navigate with the depth-based representation, with no significant drop in preferred walking speed when over-hanging obstacles were included [12]. This is in contrast to intensity-based, where a significant drop in walking speed was recorded [13].

While offering potential benefits for navigation with low resolution visual prostheses, the depth-based representation is currently limited by both noisy stereo disparity measurements, and a lack of distinction between traversable and non-traversable space. In particular, boundaries between connected surfaces become difficult to perceive as image points on either side of the boundary typically hold similar depth values. Thus, small ground surface obstacles may easily go unnoticed. Similar limitations exist under luminance-based representations, where the perception of such obstacles relies heavily on a change of intensity occuring. The clear distinction of such obstacles is essential to the safety of implantees, motivating consideration of possible augmentations to enhance the perception of non-traversable space.

In mobile robot navigation, ground-plane modelling is commonly employed to determine the traversability of the immediate space. Often this is achieved by examining range data, and inferring a *dominant* planar model via random sample consensus (RANSAC), or Hough-based voting schemes (*e.g.*, [14], [15]). Obstacles are made apparent as regions that 'disagree' with the obtained model. However, these methods do not explicitly preserve surface boundaries. Recently, we proposed a surface detection and segmentation scheme based on the examination of iso-disparity contours [16]. In this, surface boundaries are inferred as discontinuities along iso-disparity contours (and disparity gradients).

In this paper, we propose a novel scene representation for obstacle avoidance with a low resolution visual prosthesis. Using iso-disparity analysis for ground surface segmentation, we infer all traversable space in the image, from which all non-traversable surfaces are also obtained. From this data, we augment depth-based phosphene scene renderings to provide both a cleaner visualisation of the ground surface, and to enhance the distinction between traversable and non-traversable space; in particular, small ground surface obstructions.

The paper is organised in follows. Section II reviews

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

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.

iso-disparity contours, and our method for segmenting the ground surface. Section III describes our proposed augmentation of depth for navigation with prosthetic vision. Section IV presents results and Section V our conclusions.

II. GROUND SURFACE SEGMENTATION

Ground plane segmentation is applied in two steps: 1. an initial surface detection phase; and 2. a globally optimised pixel labelling phase. Before outlining these two steps, we provide a brief overview of iso-disparity contours. See [16] for further details.

A. Iso-disparity contours

Iso-disparity contours are formed from the intersection of 3D surfaces in the scene with conceptual fronto-parallel planes upon which all points project with uniform disparity. Thus, in discrete disparity space, each plane intersects the reference optical axis at the depth associated with each disparity value. In disparity space, intersections of physical surfaces with each iso-disparity plane form a level set of adjoining points in disparity space. We refer to these as iso-disparity contours.

In [16], geometric properties of iso-disparity contours across planar surfaces are exposed, and used to infer planar surfaces in the scene. Specifically, iso-disparity contours may be deemed to be co-planar if they are adjacent, near parallel, and adhere to a linear disparity gradient. If any of these conditions fail, then a surface boundary must exist. These conditions hold similarly along disparity gradient vectors. Thus, a key advantage of iso-disparity analysis is the ability to detect and preserve all connected surface boundaries.

B. Step 1: Planar surface detection

The purpose of the plane detection phase is two-fold: 1. to form a set of ground and non-ground planar models to be utilised during segmentation; and 2. to obtain priors on the labelling of pixels as ground or non-ground via observations gained from pixels along grouped iso-disparity contours. A greedy region growing algorithm is applied to group co-planar linear iso-disparity segments and detect all planar surfaces (or piece-wise planar segments) in disparity. After comparing and merging surface groupings deemed to form part of the same physical surface, the surface grouping spanning the largest image area, in the (assumed known) direction of the ground plane, is chosen to be the ground surface grouping (referred to as $\mathcal{M}_g$). All other surfaces are inserted into the non-ground set: $\mathcal{M}_n$.

C. Step 2: Ground surface labelling

Ground surface segmentation is achieved via the optimisation of an energy function defined over a regular 2D (4-connected) Markov Random Field, such that:

$$E(l) = \sum_{p \in \mathcal{P}} E_p(l) + \sum_{(p,q) \in \mathcal{N}} E_{p,q}(l_p, l_q). \quad (1)$$

where $E_p(l)$ is the unary potential defined on the assignment of label l to pixel p, $E_{p,q}(l_p, l_q)$ defines a pair-wise cost function for assigning neighbouring pixels p and q the labels l_p and l_q respectively, and $\mathcal{N}$ is the set of all neighbouring pairs. The label l signifies membership ($l =$ 'g') or not ($l =$ 'o') with the ground surface. We provide a brief description of the cost components below.

1) Unary costs: The unary potential is defined as:

$$E_p(l) = \beta_p\Big(M_p(l) + S_p(l)\Big) + \Big(1 - \beta_p\Big)R_p(l), \quad (2)$$

where the label assignment likelihood components are measured on:

- $M_p(l)$: planar model conformity with surfaces in $\mathcal{M}_l$;
- $S_p(l)$: disparity gradient direction conformity with surfaces in $\mathcal{M}_l$; and
- $R_p(l)$: intensity model conformity obtained from pixels along surface groupings in $\mathcal{M}_l$,

The weight, β_p, is used to modulate the relative contributions of disparity and non-disparity based costs by assessing the homogeneity of disparity gradient directions within a local neighbourhood $\mathcal{W}$, centred on p, such that:

$$\beta_p = \frac{1}{N} \sum_{p_i \in \mathcal{W}} \Big| \widehat{\nabla D}_{p_i} \cdot \widehat{\nabla D}_{\mu} \Big|, \quad (3)$$

where N is the total number of pixels in $\mathcal{W}$, $\widehat{\nabla D}_{p_i}$ is the disparity gradient direction at p_i, and $\widehat{\nabla D}_{\mu}$ is the mean disparity gradient direction in $\mathcal{W}$.

2) Pair-wise cost: The pair-wise smoothness term is defined as:

$$E_{p,q}(l_p, l_q) = T(l_p \neq l_q) \\ \frac{1}{\mu}\Big(\Delta_I(l_p, l_q) + \Delta_D(l_p, l_q) + \Delta_\theta(l_p, l_q)\Big),$$

where $\frac{1}{\mu}$ is a normalising scale factor, $T(.)$ is a boolean function. As is typically done, Δ_I and Δ_D reduce the cost of different pair-wise label assignments where an abrupt intensity or depth change exists. The third pair-wise component, Δ_θ is introduced to explicitly handle connected surface boundaries (which neither Δ_I or Δ_D can guarantee), such that:

$$\Delta_\theta = \exp\left(-\frac{1}{2}\Big(U_{p,q} + V_{p,q}\Big)\right), \quad (4)$$

where

$$U_{p,q} = \left(\frac{\|\nabla D_p - \nabla D_q\|}{\delta_u}\right)^2, \quad (5)$$

$$V_{p,q} = \left(\frac{\Big(1 - (\widehat{\nabla D}_p \cdot \widehat{\nabla D}_q)\Big)}{\delta_v}\right)^2, \quad (6)$$

∇D_p is the disparity gradient vector at p, and δ_u and δ_v are tunable parameters. $U_{p,q}$ measures the extent of change of disparity gradient magnitude, and $V_{p,q}$, the change of gradient direction as determined by its closest iso-disparity contour (which, by definition, is perpendicular to the local disparity gradient vector).

D. Optimisation

Optimisation of $E(l)$ is achieved using a standard implementation of the efficient min-cut/max-flow algorithm (graph-cuts) [17]. Execution time for obtaining the min-cut is less than a second on a dual core 2.7 GHz processor over 320×240 disparity images. Figures 1(a)(ii) and 2(a)(ii) show examples of the resulting segmentation.

III. Augmenting depth-based phosphene vision

The proposed depth-based augmentation modifies ground and non-ground disparity values in distinct ways. Modifications are executed as a transformation from the source disparity image, $D(p)$, to an augmented version $D^*(p)$. A phosphene rendering of $D^*(p)$ is then generated. The binary ground surface segmentation outlined previously provides a label mask for determining which modification is applied to which pixel. We outline both modifications below.

A. Ground surface disparities

All ground surface disparities are replaced by the value obtained from a planar model of best fit for the ground surface, such that:

$$D^*(p) = -\frac{1}{\gamma}\Big(\alpha p_x + \beta p_y + D_o\Big), \tag{7}$$

where $p = (p_x, p_y)$ is the image coordinate, and D_o is the disparity of the point projecting along the planar surface normal (α, β, γ). The planar surface model is obtained using all disparity values along ground surface iso-disparity contours (obtained during the surface detection phase). The restriction of seed points to iso-disparity contours acts as an outlier remover, as points along a given iso-disparity contours must have the same disparity.

B. Non-ground disparities

We emphasise the presence of ground plane obstructions by scaling up the measured disparity of non-ground surface pixels such that:

$$D^*(p) = \lambda D(p), \tag{8}$$

where p is a non-ground pixel, $D(p)$ is the measured disparity at p, and $\lambda \geq 1$ is the scaling factor.

IV. Results

We compared three scene representations: (i) intensity-based (phosphenes encode scene luminance), (ii) non-augmented depth (phosphenes encode surface proximity), and (iii) augmented depth (the proposed representation). Each phosphene is computed from regularly sampled pixels of the original image. Phosphene brightness is determined by the value, i, of the sampled pixel, (x, y), such that:

$$p[i, x, y] = be^{\sigma}\left(\frac{i}{2^{d_i} \times s^{s^{d_i} - d_o}}\right)^{\gamma} G, \tag{9}$$

where b is the brightness scale, γ is the desired gamma setting, G is a discrete Gaussian kernel centred on (x, y), d_i is the input dynamic range, and d_o is the required output dynamic range (in bits). To examine the effects of the expected reduction of dynamic range in current retinal prostheses, phosphene renderings were computed for dynamic ranges: $d_o = 6$ (6-bit, 64 brightness levels), and $d_o = 2$ (2-bit, 4 brightness levels).

Iso-disparity ground surface segmentation was executed on a sample stereo image pair and disparity data from our navigation trial environment [12], [13] (Fig 1(a)), and from a realistic outdoor scene (Fig 2(b)). The top image of both figures shows the original image, where the most prominent ground plane obstructions are a black box in the bottom-centre of Fig 1(a), and a seat and ledge to the left and right of image in Fig 2(a).

The top rows of Figures 1(b) and 2(b) show the resulting 6-bit phosphene renderings for each of the representations. In both cases, the augmented depth representation provides the clearest distinction between traversable space and ground surface obstacles. In contrast, the intensity-based representation provides few cues upon which to notice the black ground obstacle in Figure 1(b)(i), and fails to clearly delineate surface boundaries in the outdoor scene. Non-augmented depth improves the visual of the black ground obstacle (Fig 1(b)(ii)), but conveys insufficient depth variation to distinguish the seat and step from the ground surface in Figure 2(b)(ii). All obstructions are clerly evident using augmented depth (Fig 2(c)(iii)).

The bottom rows of Figures 1(b) and 2(b) show the corresponding 2-bit representations. Despite the significant reduction in dynamic range, the augmented depth still provides a clear distinction of ground surface obstacles. In both intensity and non-augmented depth, ground surface obstacles are near impossible to see.

While preliminary, these results demonstrate a clear advantage for distinguishing traversable and non-traversable space using the proposed augmented depth representation. Moreover, this advantage appears to be greater as dynamic range is reduced.

V. Conclusion

We have proposed a novel scene representation to support obstacle avoidance with a low resolution retinal prosthesis. By applying iso-disparity analysis to segment traversable and non-traversable space, we have shown how such information may be utilised to augment depth-based phosphene vision. Our results demonstrate that ground obstructions are made significantly clearer using augmented depth than with intensity or non-augmented depth, and remains informative when dynamic range is reduced. This suggests ground surface segmentation is an important operation for maintaining workable scene representations at low resolution and dynamic range.

References

[1] S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, "Simulating prosthetic vision: II. measuring functional capacity," *Vision Research*, vol. 49, no. 19, pp. 2329 – 2343, 2009.

[2] L. Hallum, G. Suaning, and N. Lovell, "Contribution to the theory of prosthetic vision," *ASAIO journal*, vol. 50, no. 4, p. 392, 2004.

[3] W. Dobelle, "Artificial vision for the blind by connecting a television camera to the visual cortex," *ASAIO journal*, vol. 46, no. 1, p. 3, 2000.

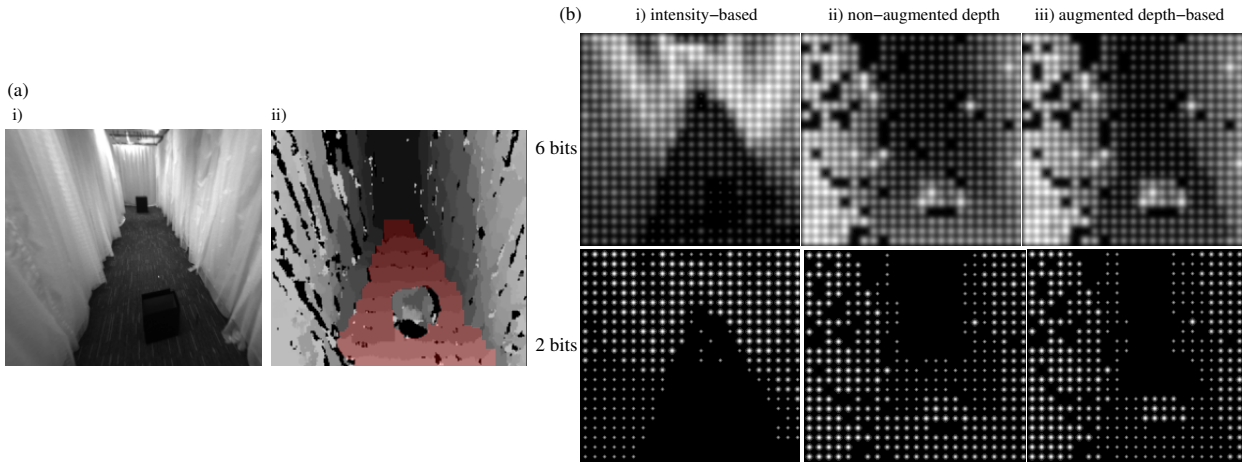

Fig. 1. Trial environment results showing: (a)(i) the original image and, (a)(ii) disparity image with ground surface segmentation marked in red; (b) top and bottom rows show results for 6 bit and 2 bit dynamic range respectively. Each row shows from left to right: (i) intensity-based phosphene rendering, (ii) a non-augmented depth-based phosphene rendering, and (iii) the proposed augmented depth-based rendering ($\lambda = 2$). Best viewed in colour.

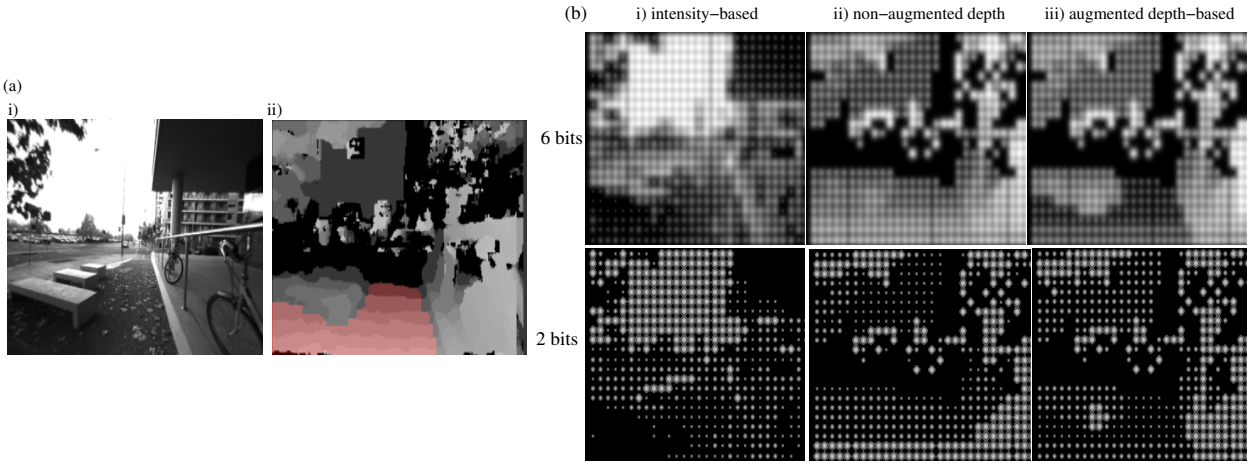

Fig. 2. Outdoor scene results. See Fig. 1 for explanation.

[4] 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.

[5] S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, "Rehabilitation regimes based upon psychophysical studies of prosthetic vision," *Journal of Neural Engineering*, vol. 6, no. 3, 2009, to appear.

[6] 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.

[7] 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.

[8] 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.

[9] 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) - Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, june 2005.

[10] 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.

[11] 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.

[12] N. Barnes, P. Lieby, H. Dennet, C. McCarthy, N. Liu, and J. G. Walker, "Mobility experiments with simulated vision and sensory substitution of depth," in *Proceedings of the Association for Research in Vision and Ophthalmology (ARVO)*, 2011.

[13] N. Barnes, P. Lieby, H. Dennet, J. G. Walker, C. McCarthy, N. Liu, and Y. Li, "Investigating the role of single-viewpoint depth data in visually-guided mobility," in *Proceedings of the Vision Sciences Society (VSS)*, 2011.

[14] P. Santana, M. Guedes, L. Correia, and J. Barata, "Saliency-Based Obstacle Detection and Ground-Plane Estimation for Off-Road Vehicles," in *Proceedings of the 7th International Conference on Computer Vision Systems (ICVS 2009)*. Springer, 2009, pp. 275–284.

[15] E. Trucco, F. Isgro, and F. Bracchi, "Plane detection in disparity space," in *Visual Information Engineering, 2003. VIE 2003. International Conference on*, July 2003, pp. 73–76.

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

[17] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary amp; region segmentation of objects in n-d images," vol. 1, 2001, pp. 105–112.