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

Image Segmentation for Enhancing Symbol Recognition in Prosthetic Vision

Lachlan Horne, Nick Barnes, Chris McCarthy, and Xuming He

***Abstract*— Current and near-term implantable prosthetic vision systems offer the potential to restore some visual function, but suffer from poor resolution and dynamic range of induced phosphenes. This can make it difficult for users of prosthetic vision systems to identify symbolic information (such as signs) except in controlled conditions. Using image segmentation techniques from computer vision, we show it is possible to improve the clarity of such symbolic information for users of prosthetic vision implants in uncontrolled conditions. We use image segmentation to automatically divide a natural image into regions, and using a fixation point controlled by the user, select a region to phosphenize. This technique improves the apparent contrast and clarity of symbolic information over traditional phosphenization approaches.**

I. INTRODUCTION

Low/impaired vision is common with prevalence rates ranging from 2.7% to 5.8% [5], [6], [7]; it is projected that in Australia these numbers will increase with the number of people with vision loss aged 40 or over rising from 575,000 to almost 801,000 by 2020 [5]. In terms of the individual, low vision is associated with impaired physical and social functioning, and reduced quality of life; in particular, mobility is often impaired. Targeting mobility and its components such as the ability to safely navigate through an environment, is one aspect of visual functioning that may lend itself to effective intervention strategies including vision-related assistive devices and the retinal implant [8].

A human with normal, or even lower than normal visual acuity, can use symbolic visual cues to navigate their environment. For example, signs showing the street address of a building allow a sighted individual to find their way around a city. Losing this ability severely restricts a person's ability to navigate and understand their surroundings. The ability to read signs, hence, would be a desirable feature for a visual prosthetic device [8], [14]. Unfortunately existing and planned visual prosthetic systems do not provide enough visual acuity to be able to identify symbols such as digits and icons readily at a distance. Higher-acuity devices can allow reading of text in controlled conditions [13] but in a navigation context, the symbols could be far away and appear in unpredictable lighting conditions.

In this paper we present an image processing system which could potentially alleviate some of these issues. The system is illustrated in Fig. 1. An image segmentation algorithm was used to automatically divide an image (a) into a set of regions (c) by appearance. We expect that taking one of these regions (d), scaling it appropriately, and displaying it using a simulated phosphene vision package (e) results in faster and more accurate symbol recognition than simply phosphenizing the entire image using a naive approach (b).

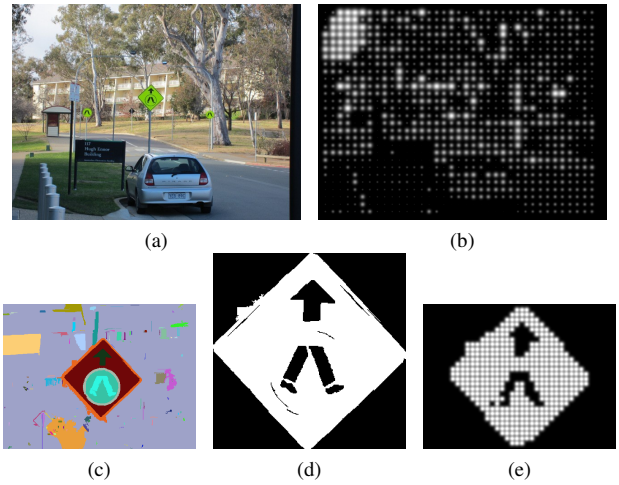


Fig. 1: (a) The original image; (b) original image phosphenized with no processing; (c) output of the first stage of segmentation; (d) result of our post-processing and fixation point region selection; (e) phosphenization of (d).

We present results showing how such segmentation might be used to improve visualisation of symbolic information. Simluated prosthetic vision images using the results of the segmentation process show symbolic and text signs in indoor and outdoor conditions show the improvement in visibility that results from this processing.

II. METHODOLOGY

A. Overview

The system requires two inputs: an image, which may be from a camera mounted to glasses worn by the user; and a fixation point, a coordinate pair on the image which may be controlled by the user or fixed. The input image is first fed into the image segmentation algorithm, which divides it into a set of regions. The region which contains the fixation point is then separated and converted to a binary image (where the region is white, and the remainder of the image is black). The binary image is then rescaled to suit the target phosphene

L. Horne, N. Barnes, C. McCarthy, and X. He are with NICTA Canberra Research Laboratory, Tower A, 7 London Circuit, Canberra ACT 2600, Locked Bag 8001, Canberra ACT 2601, Australia.

L. Horne, N. Barnes, and C. McCarthy are with the College of Engineering and Computer Science at the Australian National University (ANU), Canberra ACT 2601, Australia.

map. For demonstration, this is then fed into a phosphene vision simulator.

B. Image Segmentation

We use the method described in [1] to produce segmentations of images. A graph is generated (illustrated in Fig. 2), with one node for each pixel in the original image, and edges between adjacent pixels, with edge weights calculated as a function of pixel values. In this case, edge weights are calculated by first smoothing the image with a Gaussian convolution filter, then finding the Euclidean distance between adjacent pixel values in RGB space. This results in high weight being given to edges connecting pixels with very different color values. The image is cut into regions along boundaries with high edge weights. We added a post-processing stage to join adjacent regions of similar color together, since our desired result is a single region covering the entirety of the symbol.

The system was run with $\sigma = 0.1$, $k = 500$, $R_{min} = 20$, where σ is the scale of the smoothing filter, k is the threshold used for cutting the graph, and R_{min} is the minimum region size (regions smaller than this are merged with an adjacent region). The post-processing stage joined adjacent regions with colour difference of less than $\sqrt{5000}$ measured as Cartesian distance in RGB space. All these values were found empirically to be effective for the range of input images used.

The segmentation method produces a set of regions which group similar image pixels together. For most natural images, tens or hundreds of regions could be found in one image, most of which are irrelevant to a user of our system. To deal with this, we take only the region which contains the user's "fixation point" - which could be anywhere in the image (for example, following the user's eye gaze direction). We then use only the shape of this region to determine which phosphenes should be activated.

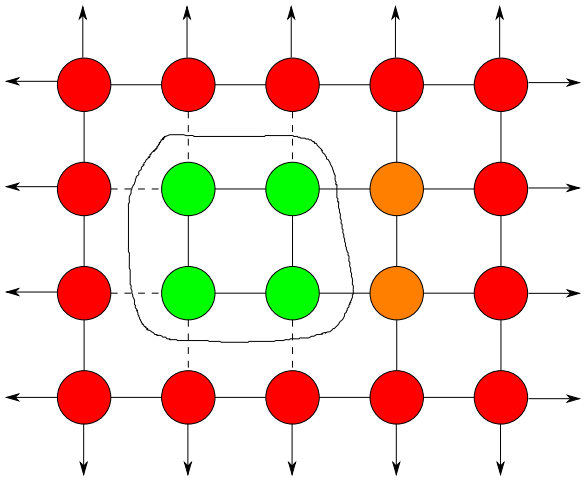


Fig. 2: A simplified example of graph-based image segmentation. A graph is formed where each node (circle) corresponds to an image pixel, and edges connect adjacent nodes. Edges are weighted by the color difference between the nodes they connect - edges with high weight are depicted as dashed lines. By removing edges along a closed path, the graph is split into multiple regions. The path is selected such that the resulting regions have a lower total edge weight - removing edges with high weight.

The fixation point is supplied to the segmentation algorithm as a parameter. Moving the fixation point can select different regions for display, and thus a user can use the fixation point to "scan" across the scene and perceive symbols with spatial context. This allows for reading simple text or groups of adjacent symbols.

Processing time varies slightly with the particular image used. All test images used were RGB, 8 bits per channel, 640×480 pixels. Processing for each completed within 190ms on an Intel Core2 Quad running at 2.66GHz, which corresponds to a frame rate of over 5 frames per second. Further optimisation could improve this significantly.

C. Simulated phosphene rendering

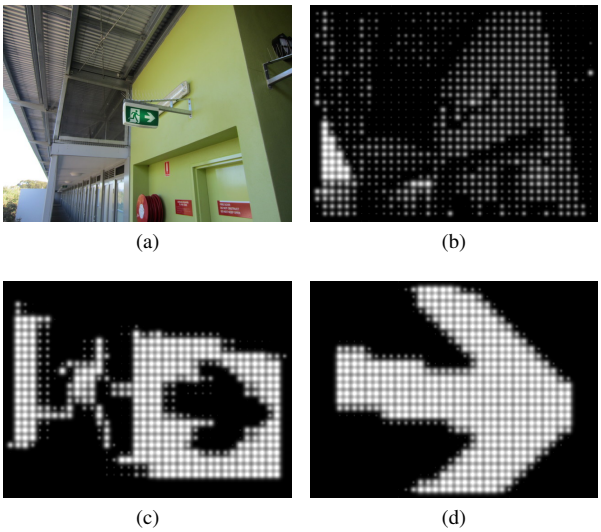
The simulated phosphene vision system described in [2] was used. Phosphenes were built as discrete Gaussian kernels using impulse sampling of the segmented region at the phosphene location, without prior filtering. Other common methods for rendering phosphenes are to filter the image prior to sampling [10] or after [11], or apply a mean or Gaussian filter over an area centered at the phosphene location [9].

Phosphene rendering was applied to both the original intensity image (for comparison) and the segmented region. Our simulated phosphene display consisted of a 35×30 rectangular grid scaled to image size. Each phosphene had a circular Gaussian profile whose center value and standard deviation is modulated by brightness at that point [12]. In addition, phosphenes sum their values when they overlap. Phosphene rendering was performed at 8 bits of dynamic range per phosphene, which is an idealised representation. There is evidence that at least 10 distinct levels of brightness can be reliably percieved by an implant recipient [15].

III. RESULTS

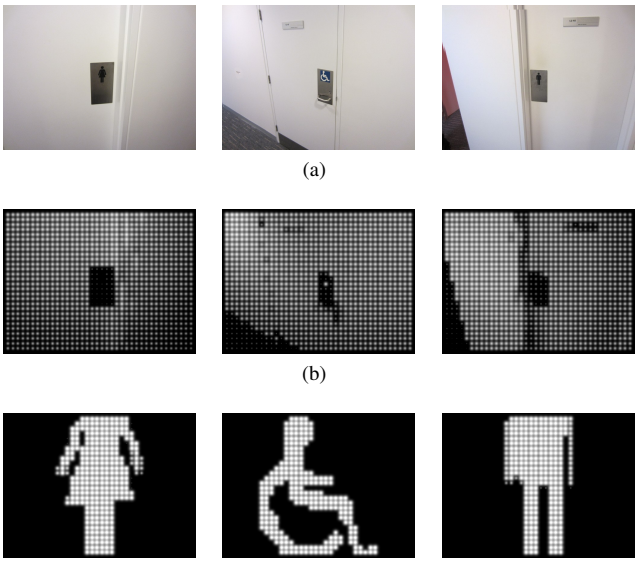
A number of example images were tested with the system and show promising results. While formal studies have not yet been performed on the utility of the system, it is clear that the original symbols can be easily recognized from the resulting phosphenizations. Significantly, the examples shown here are from a range of real-world situations, with varying lighting conditions, size of object, and contrast present in images.

An example of a typical situation where segmentation is useful is in Fig. 3. A person with impaired vision might not see the sign in (a) at all. Even with a high-acuity retinal implant, they are likely to perceive something similar to that shown in (b), which still does not clearly convey the information in the sign, or even that a sign is present. However by applying the segmentation approach described, the phosphenizations in (c) and (d) can be generated by scanning the fixation point around the image. These phosphenizations



(a) (b) (c) (d)

Fig. 3: (a) The original image; (b) original image phosphenized with no processing; (c) and (d) phosphenization of segmentation, with two different fixation points within the sign area.



(a) (b) (c)

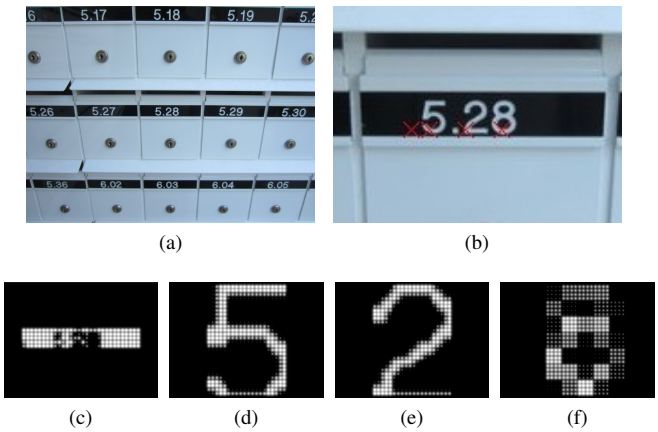
Fig. 5: (a) Original images; (b) phosphenization without processing; (c) phosphenization of segmentation.

emphasize the important information in the sign and make it readily recognizable.

Fig. 4 shows that reading simple incidental text is possible with this system. While moving the fixation point from left to right in this image, the phosphenization (c) is shown when the fixation point is over the label background, and the phosphenizations in (d), (e) and (f) are shown when the fixation point passes over the digits. The poor quality of the phosphenization in (f) is due to an image resizing error, despite this, given context, the digit depicted is recognizable.

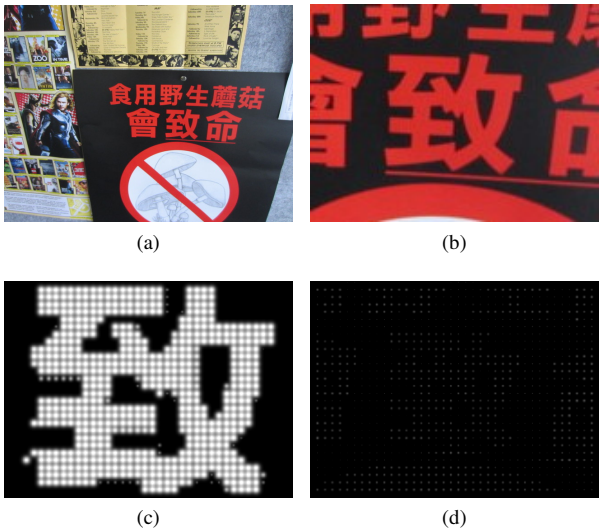
The system also allows users to discriminate between a number of different but similar symbols. Fig. 5 shows results for some common symbols. Note that in two of the three cases, the symbol is similar in shape and has poor contrast with the background, but it is relatively easy to distinguish between all three in the phosphenizations of segmentations in (c). Note also that only connected components of the fixated region are shown - i.e. the heads are not shown on the male and female symbols. In these cases moving the fixation point will allow a user to view disjoint components separately, as demonstrated in Fig. 4.

The binarization aspect of the system allows it to compensate for low dynamic range in foreseeable retinal implants. Fig. 6 shows results for a sign with poor contrast. Even after zooming the image as in (b), the symbol is not clearly represented. The segmentation approach, however, shows the symbol clearly with high contrast.



(a) (b) (c) (d) (e) (f)

Fig. 4: (a) Original image; (b) showing fixation points used (centre of each red X); (c), (d), (e) and (f) phosphenization of segmentations using fixation points, from left to right.



(a) (b) (c) (d)

Fig. 6: (a) Original image; (b) zoomed version of original image; (c) phosphenization of segmentation; (d) phosphenization of zoomed image.

IV. CONCLUSION AND FUTURE WORK

We have shown that image segmentation applied using our method is feasible and useful for a real-world prosthetic vision system, and shows promising results for symbol recognition. However, there is much exploration still to do in this space. Human trials of the system, similar to those described in [2] would be useful to confirm the usefulness of this approach. More importantly, the technique has only been tested on static images as shown here, so our next step will be to test the system with real-time video input. The segmentation algorithm should be improved with this in mind, and aim to ensure the segmentations are stable enough to be useful in this context. The system also requires further optimisation to run in real-time with an acceptable frame rate. Further development of segmentation algorithms should continue with the real-time requirements of the overall system in mind.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Paulette Lieby and Adele Scott, authors of the simulated prosthetic vision software used in this paper.

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council (ARC) through the ICT Centre of Excellence program. This research was supported by the ARC through its Special Research Initiative (SRI) in Bionic Vision Science and Technology grant to Bionic Vision Australia (BVA).

REFERENCES

[1] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, 59(2) September 2004.

[2] P. Lieby, N. Barnes, C. McCarthy, N. Liu, H. Dennett, J.G. Walker, V. Botea, and A.F. Scott. Substituting depth for intensity and real-time phosphene rendering: Visual navigation under low vision conditions. EMBC, 2011.

[3] N. Barnes, P. Lieby, H. Dennet, C. McCarthy, N. Liu, and J.G. Walker. Mobility experiments with simulated vision and sensory substitution of depth. ARVO, 2011.

[4] 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. VSS, 2011.

[5] Access Economics Pty Limited. Clear focus: The economic impact of vision loss in australia in 2009, 2010.

[6] H. Taylor, J. Keeffe, H. Vu, J. Wang, E. Rochtchina, and M.L. Pezzullo. Vision loss in australia. Medical Journal of Australia, 182:565568, 2005.

[7] T. Y. Wong, E. W. Chong, W.-L. Wong, M. Rosman, T. Aung, J.-L. Loo, et al. Prevalence and causes of low vision and blindness in an urban malay population: The singapore malay eye study. Arch Ophthalmol, 126(8):10911099, 2008.

[8] J.E. Keeffe, K.L. Francis, C.D. Luu, N. Barnes, E.L. Lamoureaux, and R.H. Guymer. Expectations of a visual prosthesis: perspectives from people with impaired vision. ARVO, 2010.

[9] S.C. Chen, G.J. Suaning, J.W. Morley, and N.H. Lovell. Simulating prosthetic vision: I. Visual models of phosphenes. Vision Research, 2009.

[10] S.C. Chen, L.E. Hallum, N.H. Lovell, and G.J. Suaning. Visual acuity measurement of prosthetic vision: a virtual-reality simulation study. J. Neural Eng., 2:S135S145, 2005.

[11] J. Dowling, W. Boles, and A. Maeder. Simulated artificial human vision: The effects of spatial resolution and frame rate on mobility. In Active Media Technology, number 138, pages 138143, 2006.

[12] M. Vurro, G. Baselli, F. Orabona, and G. Sandini. Simulation and assessment of bioinspired visual processing system for epi-retinal prostheses. In Engineering in Medicine and Biology Society, 2006. EMBS 06. 28th Annual International Conference of the IEEE, 2006.

[13] E. Zrenner, K.U. Bartz-Schmidt, H. Benav, D. Besch, A. Bruckmann, V. 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. In Proc. R. Soc. B, November 3, 2010.

[14] J. Sommerhalder, How to Restore Reading With Visual Prostheses. In Visual Prosthesis and Opthalmic Devices, Opthalmology Research, Humana Press 2007

[15] M.S. Humayun, J.D. Weiland, G.Y. Fujii, R. Greenberg, R. Williamson, J. Little, B. Mech, V. Cimmarusti, G. Van Boemel, G. Dagnelie, E. de Juan Jr., Visual perception in a blind subject with a chronic microelectronic retinal prosthesis, Vision Research, Volume 43, Issue 24, November 2003, Pages 2573-2581, ISSN 0042-6989, 10.1016/S0042-6989(03)00457-7.