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

# Substituting depth for intensity and real-time phosphene rendering: Visual navigation under low vision conditions

Paulette Lieby, Nick Barnes, Chris McCarthy, Nianjun Liu,
Hugh Dennett, Janine G. Walker, Viorica Botea, and Adele F. Scott

*Abstract*— **Navigation and way finding including obstacle avoidance is difficult when visual perception is limited to low resolution, such as is currently available on a bionic eye. Depth visualisation may be a suitable alternative. Such an approach can be evaluated using simulated phosphenes with a wearable mobile virtual reality kit.**

**In this paper, we present two novel approaches: (i) an implementation of depth visualisation; and, (ii) novel methods for rapid rendering of simulated phosphenes with an empirical comparison between them. Our new software-based method for simulated phosphene rendering shows large speed improvements, facilitating the display in real-time of a large number of phosphenes with size and brightness dependent on pixel intensity, and with customised output dynamic range.**

**Further, we describe the protocol, navigation environment and system used for visual navigation experiments to evaluate the use of depth on low resolution simulations of a bionic eye perceptual experience. Results for these experiments show that a depth-based representation is effective for navigation, and shows significant advantages over intensity-based approaches when overhanging obstacles are present. The results of the experiments were reported in [1], [2].**

## I. INTRODUCTION

Low/impaired vision is common with prevalence rates ranging from 2.7 to 5.8% [3], [4], [5]; 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 [3]. 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 [6].

In normal human vision, depth cues come from many sources including stereo disparity, optical flow, scene understanding and motion parallax. These cues have a role in enabling visual navigation, and the direct availability of depth cues may be important for low resolution visual displays, such as used in a bionic eye. In this paper we describe the implementation of depth visualisation in such displays. We also describe the technical setup and protocol of mobility experiments comparing the efficacy of depth-based visualisation versus intensity-based visualisation.

Simulated prosthetic vision (SPV) was used in these experiments. A *phosphene* is the visual percept experienced by a subject when one site of the visual pathway is under electrical stimulation. Although phosphenes have been reported as presenting a variety of shapes, they are commonly represented as dot-like [7]. SPV is typically implemented as a collection of (simulated) phosphenes [7], [8] arranged on a regular or hexagonal grid. While SPV is an idealised representation of the prosthetic experience, it is well suited as a first step to evaluating the functionality of visual representations.

We trialled two visual representations: the first consisted entirely of disparity information directly available across the whole of the visual field, while the second was intensity-based and depth poor. Resolution was low ($35 \times 30$ visual elements/phosphenes) so that there was little available in terms of depth cues from motion parallax, shadowing, or linear perspective.

Participants were blindfolded, and access to visual information was via a head mounted display (HMD) connected to head-mounted cameras. Disparity and intensity images were captured directly from the cameras. In the disparity image the pixel value is relative to the location mapped by the pixel: the brighter the pixel, the closer the location. Phosphenes were built as discrete Gaussian kernels using impulse sampling at the phosphene location, without prior filtering. Other common methods for rendering phosphenes are to filter the image prior to sampling [8] or after [9], or apply a mean or Gaussian filter over an area centered at the phosphene location [7].

There have been many studies assessing SPV for navigation. Some studies use static images [10] or simulate phosphenised vision by covering the camera with a mask [11]. A few studies implement close to real-time simulation [12], [9]. For experiments using SPV in an interactive setting, such as navigation trials, fast rendering of the simulated phosphenes is critical. In this paper, we present a purely software-based method for rendering phosphenes to support a broad variety of possible scenarios, including from very few up to tens of thousands of phosphenes in a single image, a varying output dynamic range (possibly varying across different phosphenes in a single image), and different phosphene sizes. We compare our approach with two Gaussian convolution-based approaches that use some fast tricks in terms of implementation and a new method for Fast Gaussian Approximation [13]. We show that a lookup table

P. Lieby, N. Barnes, C. McCarthy, N. Liu, V. Botea, and A. Scott are with NICTA Canberra Research Laboratory, Tower A, 7 London Circuit, Canberra ACT 2600, Locked Bag 8001, Canberra ACT 2601, Australia.

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

H. Dennett is with the Department of Psychology at ANU, and J.G. Walker is with the Center for Mental Health Research, also at ANU.

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

offers best (real-time) performance up to 5,000 phosphenes over all output dynamic range values.

## II. METHODS

### A. The study

The study was a pilot to inform a larger forthcoming study; it had a 2 × 2 repeated measures design. Specifically the design reflects the two levels of depth perception (depth rich, depth deprived) and two levels of navigation task difficulty (with or without overhanging obstacles). Human ethics approval has been granted by the Human Research Ethics at the Australian National University, Canberra, Australia.

### B. Participants

Four participants (three females and one male), aged 21-24, with objectively assessed (by testing visual acuity and contrast sensitivity) normal or corrected-to-normal vision and no mobility impairment were recruited.

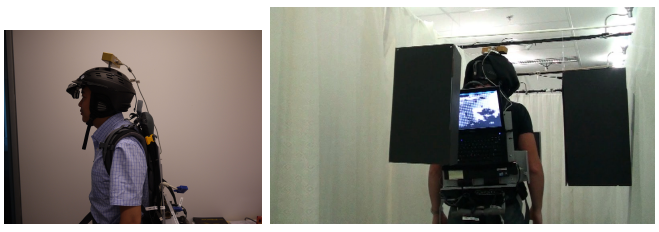
### C. The trial environment

The navigation trials took place in an indoor modular maze comprising 3 × 6 cubicles of 1.5m × 1.5m. The cubicles' partitions were made of cloth curtains that were moved according to the trail course to be undertaken; fabric drops were 'sewn' together using Velcro strips. For some traversals overhanging obstacles were added; they were made of fabric. The fabric drops were white, while the floor and the obstacles were dark grey and black respectively, see Figure 1 (b). The contrasted colour scheme provides well-suited conditions for differentiating intensity-based information at low resolution (see studies like [11] for example). We added textured lace to walls and obstacles to enable computation of the depth-based representation.

### D. The physical visual system

A modified skate-board helmet was used to hold a stereo camera rig (Point Gray Bumblebee) together with a HMD, the eMagin Z800 3D Visor. The helmet shows great stability during mobility trials. We made a blindfold out of fabric to achieve effective blocking of any visual cues beside those displayed by the HMD.

Visual processing was performed on a laptop mounted on a converted baby carrier worn by the participant. The laptop monitor was used by experimenters to monitor what was effectively seen by the participant. Figure 1 shows the resulting set up.



(a) (b)

Fig. 1. (a) the physical visual system and (b) navigating through the maze.

### E. Outcome Measures

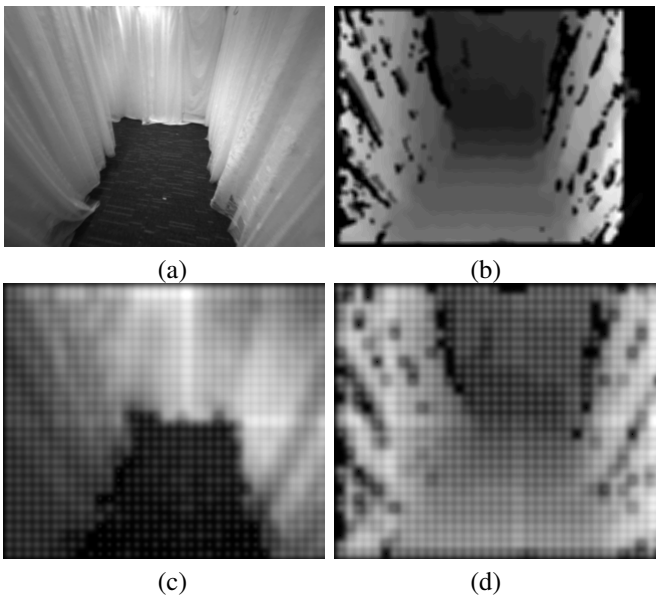
Outcome measures were aspects of human navigation and included percentage of preferred walking speed (PWS), and number of contacts with objects.

*1) Percentage of preferred walking speed (PPWS):* PPWS is the ratio of mobility course speed to preferred walking speed, expressed as a percentage [14]. The mobility course speed is the speed at which participants navigate the trial environment under experimental conditions; preferred walking speed is the speed at which participants navigate under normal vision conditions (i.e., without vision processing).

*2) Errors (contacts with walls and obstacles):* Errors are defined as contact with walls and obstacles. Each trial was monitored by an experimenter who recorded the number of collisions with walls and obstacles in the test environment.

### F. Simulated phosphene rendering

Phosphene rendering was applied either to the intensity image or to the depth image; the images had a resolution of 320 × 240 pixels. 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 [15]. In addition, phosphenes sum their values when they overlap. For the pilot study phosphene rendering was performed at 8 bits which again is an idealised representation. It is generally assumed that maximum neuronal discrimination of electrical stimulation is closer to a 3 bit rendering. Figure 2 shows the phosphenised rendering of the intensity image (a) and the depth image (b).



(a) (b)

(c) (d)

Fig. 2. (a): the original image; (b) the depth image (here normalised for easy viewing); (c) the phosphenised intensity image; (c) the phosphenised depth image.

*1) A real-time rendering method: In-place kernel method:* Let $d_i$ and $d_o$ be the input and output dynamic range respectively. In our case $d_i = 8$. We sample the image at the phosphene location $(x, y)$, and using the intensity $i$ at that

location (without additional filtering), we create a discrete Gaussian kernel $G$ of size $w \times w$ centered at $(x, y)$ with standard deviation ($\lfloor \cdot \rfloor$ denotes the rounding function)

$$\sigma = I^{\gamma}, \tag{1}$$

$$I = \left\lfloor \left\lfloor i \times \frac{2^{d_o} - 1}{2^{d_i} - 1} \right\rfloor \times \frac{255}{2^{d_o} - 1} \right\rfloor, \tag{2}$$

$$\text{and } w = \lfloor 6\sigma \rfloor. \tag{3}$$

Here $\gamma$ is the gamma control parameter. With $b$ a brightness parameter the final phosphene value at $(x, y)$ is set to

$$p[i, x, y] = \frac{255}{I_c} \quad b \quad e^{\sigma} \quad \left(\frac{I}{255}\right)^{\gamma} G[x, y]. \tag{4}$$

where $I_c$ is a normalising constant. This ensures that both the size and the brightness of the phosphene are dependent on $i$. The setting of the standard deviation $\sigma$ to $I^{\gamma}$ has been chosen as it gives good empirical results in modulating phosphene size with intensity.

Constructing a lookup table of $p[i]$ matrices for all possible values of $i$ ensures fast computation: the final phosphene image is obtained by simply summing up the $p[i, x, y]$ matrices for all phosphene locations $(x, y)$. Note that our in-place kernel method allows for each phosphene $p$ being rendered according to individualised parameters $\gamma_p$ and $b_p$. With careful construction of the lookup table with respect to the $\gamma_p$ and $b_p$ parameters it is possible to minimise the size of the table and thus the amount of overhead.

*2) A convolution approach:* We compared our approach with the traditional convolution approach whereby the whole image is convolved with a Gaussian. First we create $2^{d_o}$ blank images $\mathcal{I}_n$, $n = 1, \ldots, 2^{d_o}$, of same size as the original image $\mathcal{I}$.

We sample $\mathcal{I}$ at the phosphene location $(x, y)$ (again, without additional filtering), giving an intensity $i$. For $n$ such that $i \in [(n-1)2^{d_i - d_o}, n2^{d_i - d_o} - 1]$ the value at $(x, y)$ in image $\mathcal{I}_n$ is set to $\frac{1}{I_c}$ $b$ $e^{\sigma}$ $\left(\frac{I}{255}\right)^{\gamma}$ where $I$, $\gamma$, $b$, $I_c$, and $\sigma = I^{\gamma}$ are as defined above. Each of the $\mathcal{I}_n$ images is then convolved with a Gaussian with standard variation $\sigma$. Adding the $\mathcal{I}_n$ images and mapping the pixel values into the $[0 \ldots 255]$ range gives the final phosphene image. Bar the normalisation inherent in the convolution of the $\mathcal{I}_n$ images, the final phosphene image is equivalent to the one obtained by the in-place kernel method. We opted for a layered approach in order to render phosphenes with a size varying with intensity.

We implemented the convolution of each image $\mathcal{I}_n$ (i) using the highly optimized openCV `cvSmooth` function with a Gaussian filter (implemented using FFT), and (ii) using a fast Gaussian approximation based on an averaging filter as described in [13].

*3) Fast Gaussian approximation:* Fast Gaussian approximation is implemented as a sequence of averaging filters using integral images. One averaging filter of width $w$ has standard deviation of $\sqrt{\frac{w^2-1}{12}}$; $n$ averagings with the same filter will have a standard deviation of $\sqrt{\frac{nw^2-n}{12}}$. Then the ideal width $w_i$ of the averaging filter approximating a Gaussian filter of standard deviation $\sigma$ will be $w_i = \sqrt{\frac{12\sigma^2}{n} + 1}$. We require that $w_i$ be an odd integer; the solution is to find the two odd integers $w_l$ and $w_u = w_l + 2$ with $w_l <= w_i$ and $w_u >= w_i$ together with the integer $m$ so that $m$ averagings with a filter of width $w_l$ followed by $n - m$ averagings with a filter of width $w_u$ approximates the original Gaussian filter of standard deviation $\sigma$. For $n = 5$, the accuracy of the achieved standard deviation is $\pm$ 0.1673 [13].

## III. Results

Since the Fast Gaussian approximation is a sequence of averaging filters it will be constant in the size of the window or phosphene size $w$, itself dependent on $\gamma$, see (3) and (1). This is opposed to the openCV Gaussian `cvSmooth` function, which is exponential in $\gamma$, that is $w$; in-place kernel also depends exponentially on $\gamma$. Both convolution methods are constant in the number of phosphenes, while in-place kernel is linear. Finally the two convolution approaches are exponential in the number of layers, that is in the dynamic range value, while in-place kernel is constant.

To compare the three methods we proceed as follows: for a given resolution of $320 \times 240$ pixels and for a given number of phosphenes $n$, we compute a maximum phosphene size $w = 3d$ where $d$ is the inter-phosphene distance in pixels when $n$ phosphenes are uniformly placed in a grid on the whole image. A value of $w$ larger than $3d$ would not be realistic as rendered phosphenes would entirely overlap their closest neighbours. In other words $w$ is a function of $n$ (and of the image resolution). For given $n$ and $w$ dependent on $n$ and a dynamic range of 3 bits, we time the three methods; Figure 3 shows the results.

In-place kernel outperforms the convolutions methods for $n$ up to 5,000. In Figure 3 we note that the openCV Gaussian method outperforms the Fast Gaussian approximation for values of $n$ larger than 200: this is easily explained by the fact that $w$ gets smaller as $n$ increases. Fast Gaussian approximation has real-time performance for any number of phosphenes; openCV Gaussian achieves real-time performance for $n > 50$ (Figure 3(b)).

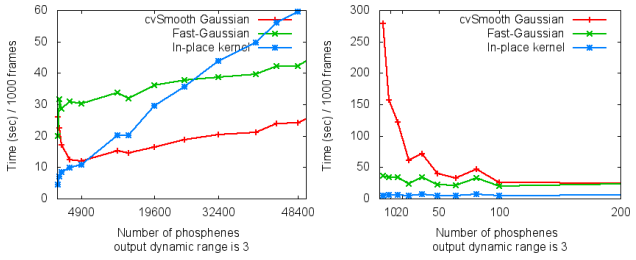
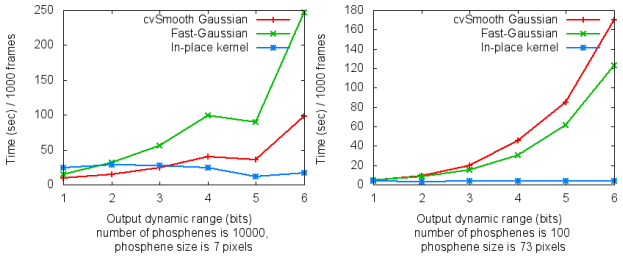



Fig. 3. Comparing in-place kernel with the two convolution methods, openCV Gaussian `cvSmooth` and Fast Gaussian approximation for varying number of phosphenes, a maximum phosphene size dependent on the number of phosphenes, and an output dynamic range of 3bits (see text).

For larger dynamic ranges we compare the performances of the three methods for $n = 1000$ and $n = 100$; as before the phosphene size $w$ is a function of $n$. Figure 4 shows the results: in-place kernel outperforms the convolution methods,

and is real-time and nearly constant for all dynamic ranges. Again compare Figure 4(a) and Figure 4(b) for the differing performance of Fast Gaussian approximation and openCV Gaussian with respect to phosphene size.




(a) (b)

Fig. 4. Comparing in-place kernel with the two convolution methods, openCV Gaussian `cvSmooth` and Fast Gaussian approximation for varying output dynamic range, for $n = 100$ (a), and $n = 100$ (b).

We conclude that for simulations typical of low vision conditions (e.g. less than 5,000 phosphenes) in-place kernel outperforms the convolution methods over a variety of dynamic range values while uniquely providing the possibility for individual phosphene parameterisation. This option is not available in the openCV Gaussian method which could be considered when the number of phosphenes is larger than 5,000 and the output dynamic range is 3, a value closer to the actual behaviour of discriminative electrical neuronal stimulation. Note also that in-place kernel is real-time for a large number of phosphenes (up to 30,000) over the whole dynamic range.

As Fast Gaussian approximation is constant in phosphene size it could be considered if larger image resolution were necessary; the other methods would do poorly as they are exponential in $w$ – large Gaussians are computationally expensive.

Real-time phosphene rendering enabled us to successfully run human mobility pilot trials under low vision conditions. Depth and intensity images, both obtained from the Point Gray Bumblebee camera, were phosphenised. Participants completed various navigation tasks under a randomly allocated visual representation (either depth or intensity).

Depth- and intensity-based representations were effective, facilitating visually guided navigation (PPWS was larger than 50% in all cases –with and without obstacles, $n = 54$, $p < 0.05$) [1]. When no obstacles were present, intensity facilitated significantly faster navigation than depth ($p < 0.001$). However, when overhanging obstacles were present, PPWS was significantly reduced for intensity ($p = 0.03$)); this was not the case for the depth-based representation. However the difference between the difference of PPWS for depth and the difference of PPWS for intensity was significant ($p = 0.01$). Therefore a depth-based representation shows significant advantage when overhanging obstacles are present in the navigation environment [2].

## IV. Conclusion and Future Work

In this paper we have described the technical implementation and protocol for experiments in human mobility under low vision conditions. We have also introduced and described a software-based real-time phosphene rendering method suitable for a large number number of arbitrarily placed phosphenes, and an arbitrary dynamic range. As a next step, we plan to investigate visual representations where rendering is customised for each phosphene, thus fusing image processing and rendering.

## V. Acknowledgments

Thanks are due to Hongdong Li for the layered convolution approach. We also thank the research assistants Joanna Cheng, Hanxi Li, and Adnan Shah who helped in running the experiments.

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] 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.

[2] 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.

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

[4] H. Taylor, J. Keeffe, H. Vu, J. Wang, E. Rochtchina, and M.L. Pezzullo. Vision loss in australia. *Medical Journal of Australia*, 182:565–568, 2005.

[5] T. Y. Wong, E. W. Chong, W.-L. Wong, M. Rosman, T. Aung, J.-L. Loo, and et al. Prevalence and causes of low vision and blindness in an urban malay population: The singapore malay eye study. *Arch Ophthalmol*, 126(8):1091–1099, 2008.

[6] 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.

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

[8] 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:S135–S145, 2005.

[9] 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 138–143, 2006.

[10] J. Dowling, A. Maeder, and W. Boles. Mobility enhancement and assessment for a visual prosthesis. In *SPIE Medical Imaging 2004: Physiology, Function, and Structure from Medical Images*, 2004.

[11] K. Cha, K.W. Horch, and R.A. Normann. Mobility performance with a pixelized vision system. *Vision Res.*, 32(7):1367–1372, 1992.

[12] G. Dagnelie, P. Keane, V. Narla, L. Yang, J. Weiland, and M. Humayun. Real and virtual mobility performance in simulated prosthetic vision. *J. Neural Eng.*, 4:S92–S101, 2007.

[13] P. Kovesi. Fast almost gaussian filtering. In *Digital Image Computing, Techniques and Applications, 2010. 12th International Conference on Digital Image Computing*, pages 121–125, 2010.

[14] D.D. Clark-Carter, A.D. Heyes, and C.I. Howarth. The efficiency and walking speed of visually impaired people. *Ergonomics*, 29(6):779–789, 1986.

[15] 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.