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

Text image processing for visual prostheses

Song Wang, Yi Li and Nick Barnes

Abstract—**Retinal diseases are leading causes of severe vision loss and blindness throughout the world. Visual prosthetics has been demonstrated to be an effective therapy to partially restore vision. Reading is one of the most important functional abilities derived from vision. Therefore, we propose a text image processing strategy for visual prostheses. Text information is firstly detected and recognized from images acquired from a camera, and then recognized text is represented by simplified characters so as to be displayed in low-resolution phosphene vision. In this paper, a BP neural network is created to recognize text information in the images acquired from a high-resolution camera. The recognized text is represented using simplified characters with the resolution of 5×7 pixels. In order to mimic the visual percepts and to evaluate potential benefits of this proposed text processing strategy, a simulation model of prosthetic vision is created based on the reported visual characteristics of elicited phosphenes. Simulated prosthetic vision using 25×25 distorted phosphene array covering four letters shows phosphene letters that can be read readily. Compared to the displayed phosphene letters without this strategy, the contours of all the phosphene letters processed by this strategy were more intact and clearer. These results demonstrate benefits of this proposed strategy which is aimed to provide better reading experience for blind patients using prosthetic vision.**

I. Introduction

Retinal diseases such as retinitis pigmentosa (RP) and age-related macular degeneration (AMD) are two of the leading causes of substantial vision loss or blindness worldwide [1, 2]. Visual prosthetics has been demonstrated to be an effective therapy to partially restore visual perception in human clinical trials [3-5]. Retinal prostheses use a camera to detect light, convert light energy into an electrical signal, and deliver the electrical signal to the retinal neurons through implanted microelectrode array to elicit vision [4]. The nature of prosthetic vision is constructed from electrically-elicited percepts in the visual field which are called “phosphenes”. The possibility of restoring partial vision relies on multiple simultaneously elicited phosphenes, which is the foundation for current clinical trials and is the general assumption made in the studies of image processing for visual prostheses [6].

Human clinical trials have reported the visual characteristics of elicited phosphenes, including shape, brightness, size and position, as well as the percepts in response to multiple concurrent activation. As the patterns of neural activity elicited by electrical stimulation of the retina will depend on stimulating current strength and on the distance between the electrode and the neural target, the elicited phosphenes can be predicted to be neither of regular shape nor constant luminosity [7]. The common observed shapes of the phosphenes elicited at the retina are approximately round or oval, curved, straight short lines, or in the form of wedges [5, 8, 9]. Also, Dagnelie et al., who had first-hand contact with subjects, indicated that the elicited phosphenes did not resemble sharp-edged round dots [10]. Phosphenes are generally bright and readily visible, and the identified brightness ratings were reported to have five to ten levels [5, 11]. The size of phosphene varies greatly from a punctuate spot of light, about 0.1 degree of visual angle, to as large as a football at arm’s length, about 25 degrees of visual angle, while most phosphenes subtended around 0.5 to 2 degrees of visual angle [5, 9, 12, 13]. The positions of perceived phosphenes in general matched the positions of stimulating electrodes on the retina and a subject was able to identify which electrodes were activated based on the positions of the phosphenes [5]. Human trials involving the use of compound electrodes have shown that perception of primitive shapes formed by multiple phosphenes is feasible [14]. Also, Dorn et al. reported that one subject using an epi-retinal prosthesis with 10×6 electrodes could perceive complex shapes with intersecting lines, such as an “H”, a triangle, a “T”, and several parallel lines [15]. Based on these visual characteristics of elicited phosphenes, simulated prosthetic vision is created in this paper. It is used to mimic the visual percepts and to test the proposed text image processing strategy on normal subjects. The information about the potential benefits could be obtained without using the invasive visual prostheses. Also, the effects of single parameters may be varied over a full range for study and experiments are easy to repeat.

Research on simulation of prosthetic vision has investigated the minimum requirements for visual prostheses to restore reading abilities. Cha et al. used a pixelized vision system to simulate artificial vision in normal subjects. Their results showed that a 25×25 pixel array covering four letters of text is sufficient to provide reading rates close to 170 words/min using scrolled text, and close to 100 words/min using fixed text [16]. Dagnelie et al. indicated that reading performance deteriorated when less than four letters were displayed [7, 10]. Zhao et al. indicated that distortion of pixelized array, dropout percentage, and pixel size variability had a significant impact on the recognition of pixelized Chinese characters [6]. However, most of the previous studies were based on the assumption of regular round or square phosphene shape. Also, recent literature do assume that phosphene size and brightness were mutually independent, however, both can increase with an increase of stimulation frequency [9]. Our text image processing strategy is tested on simulated prosthetic vision consisting of 25×25 pixels covering four letters. We extend on previous prosthetic vision

This work is part of the Bionic Vision Australia (BVA) supported by Australian Research Council grant.

Song Wang, Yi Li and Nick Barnes are with the Research School of Engineering, Australian National Univeristy, Canberra, Australia, as well as with NICTA Canberra Research Laboratory, Canberra, Australia (corresponding author: nick.barnes@nicta.com.au).

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

simulation experiments with text by including phosphenes of various shapes on a distorted grid and adding the constraint of phosphene brightness and size.

Recreational reading was reported among the most important daily living activities by patients with low-vision [17]. Also, Humayun et al. who have much interaction with blind patients indicated that reading is thought by the blind patients as one of the most important visual functions among mobility without a cane, face recognition, and reading [18]. Hence, helping subjects to read is important for a visual prosthesis. For reading a document, a camera acquires images of the document, which are then resized to low resolution images. The information of each pixel in the downgraded images is used to control the stimulation parameters of each channel which conveys current to each implanted electrode and elicits a phosphene. However, the text may appear too small in the captured images, and subjects may not be able to read the text. Also, using a zoom-in function is not convenient for subjects, because the sizes of the headers and the main body of magazines or newspaper are quite different, which requires subjects to always zoom in or out the text. In addition, as the images acquired by camera are resized to low-resolution images, useful information will inevitably be lost, as shown in Figure 1. Hence, a text image processing strategy is necessary.

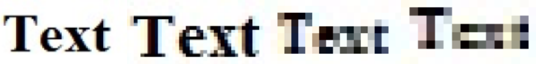


Figure 1. The word “Text” at five degrees of pixelization. (a): 80×40 pixels. (b) 40×20 pixels. (c) 20×10 pixels. (d) 16×8 pixels.

In this paper, we firstly create a simulation model of prosthetic vision according to the visual characteristics of elicited phosphenes, which is used to mimic the visual percepts and to evaluate potential benefits of the proposed text image processing strategy. A feed-forward BP neural network is created to recognize characters in the image acquired by a camera. The recognized letters are represented by simplified 5×7 pixels characters, and the text information is displayed in the simulated prosthetic vision. Finally this processing strategy is verified to be effective in simulated phosphene images; all the text information can be recognized correctly and displayed intact, and appear clearly in the distorted 25×25 phosphene array covering four letters. Also, the results of proposed strategy are compared to those without using this strategy, demonstrating that this strategy could better display text information using limited electrodes.

II. Methods

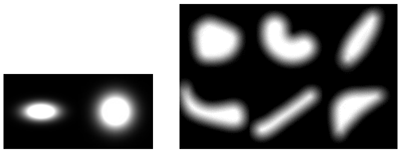
A. Simulated prosthetic vision

1. Phosphene shapes

As most elicited phosphenes were approximately round or oval, and did not resemble sharp-edged profiles, the profiles of round and oval phosphenes are created using a 2D Gaussian function expressed as,

$$f(x,y) = A \cdot e^{-(\frac{(x-x_0)^2}{2\sigma_x^2}+\frac{(y-y_0)^2}{2\sigma_y^2})}$$

where $f(x,y)$ is the value of pixel (x,y), A is the amplitude controling the center pixel value, (x_0, y_0) is the center position of a phosphene, σ_x and σ_y adjust the spreads in x-axis and y-axis, respectively. The created round and oval phosphenes are shown in Figure 2(a). The irregular shaped phosphenes, including the curved and straight short lines, and wedges, are created using subjects’ drawings of perceived phosphenes [9, 19-21] which are then binarized and blurred using a circular averaging filter as the correlation kernel. The created irregular shaped phosphenes are shown in Figure 2(b).



(a) (b)

Figure 2. Simulated shapes of phosphenes. (a): Round and oval phosphenes. (b): Irregular shaped phosphenes

2. Phosphene brightness

As subjects identified five to ten levels of phosphene brightness, the brightness scale is set from zero to ten with zero representing no perception and ten being the brightest.

3. Phosphene sizes

Nanduri et al. constructed a quantitative model successfully replicating the general findings of stimulation frequency and amplitude modulation on the size and brightness of elicited phosphenes [9]. As stimulation frequency has smaller effect on the phosphene size than the stimulation amplitude, their frequency modulation is used and the data of size and brightness is obtained with frequency ranging from 10Hz to 120Hz from this model. Using the linear regression, the quantitative relationship of the brightness factor and size factor is expressed as,

$$S(i,j) = 1.278B(i,j) + 1.043$$

where $B(i,j)$ is the brightness factor of a phosphene (i,j), ranging [0, 1.0] with 0 representing no perception and 1.0 the brightest, and $S(i,j)$ is the size factor of the phosphene (i,j). The simulated phosphenes with size and brightness properties are shown in Figure 3, which ignores the modulation of phosphene shapes.

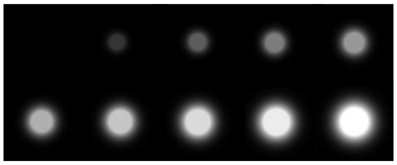


Figure 3. Simulated phoshenes with both size and brightness properties

4. Phosphene grid

Although phosphenes appear distorted with respect to the positions of the regular electrodes, the positions of perceived phosphenes in general matched the positions of the stimulating electrodes on the retina. Zhao et al. proposed Gaussian distribution to simulated this distortion [6]. Hence, two independent normal distributions, with mean zero and standard deviation (SD) fifteen percent of the center to center distance between two neighboring phosphenes, are used to mimic this distortion, which create the deviations of each phosphene in horizontal and vertical direction from a regular phosphene grid. An example of a 4×4 distorted phosphene map is shown in Figure 4.

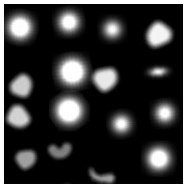


Figure 4. An example of a 4×4 distorted phosphene map

B. Image processing and text recognition

1. Image pre-processing

Pre-processing includes three parts, binarization, morphological operation, and segmentation. Firstly the high resolution text images acquired by the camera are converted to grayscale images, and then the values of all pixels in the grayscale images with luminance greater than a threshold value are replaced by 1 and all the other pixels are replaced with the value 0. By checking the connectivity within the binary images, all the boundaries are identified. Then these boundaries are dilated with a square structuring element, and all holes, sets of isolated background pixels, are filled with the value 1. Next, through checking its horizontal projection, text is split into individual lines. Then in these split images, all the 8-connected areas are labeled, and the bounding box of each letter is created according to the labels. For the letters that are not connected, i.e. i and j, through checking the centroid positions and areas of every two neighboring connected regions, these letters could be identified and then their bounding boxes could be created. According to the created bounding boxes, each letter is segmented. In these segmented images, the rows and columns with all pixel values 1 are deleted. Hence each letter is cropped sharp to its border for feature extraction.

2. Feature extraction

All segmented letter images are resized to the resolution of 50×70, and then each letter image is divided into 5×7 sub-images with each sub-image having 10×10 pixels. We apply the approach of Bokser who computed the percentage of black pixels in each zone for classification [22], the ratios of the pixels in black over the pixels in white of each sub-image are concatenated into one vector having 35 values, which is used as the feature and is fed into the neural network with 35 input neurons for pattern recognition.

3. Pattern recognition

A feed-forward back propagation neural network is constructed for pattern recognition. This neural network consists of two layers, with 35 nodes in the input layer, 14 nodes in the hidden layer and 52 nodes in the output layer. The transfer function of each node in the hidden layer and output layer uses log-sigmoid function. This neural network is trained using scaled conjugate gradient (SCG) algorithm which performs pattern recognition fast and accurately. The training set consists of 24 identical groups of Latin letters in the style of Times New Roman and each group is consisted of 26 upper and 26 lower letters.

4. Text representation

Each recognized letter is represented by 5×7 pixels character code, as shown in the Figure 5. As the size of phosphene array is 25×25 phosphenes, four letters could be displayed at the same time.

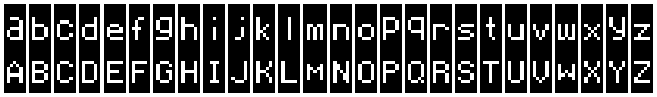
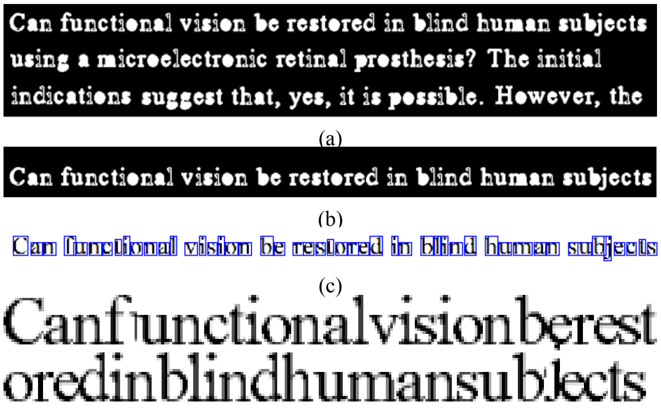

a b c d e f g h i j k l m n o p q r s t u v w x y z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Figure 5. Defined characters

III. Results

The simulated prosthetic vision image processing and text recognition are implemented in the MATLAB environment. The neural network is trained with a training dataset of 24 groups of all upper and lower Latin letters. One paragraph is captured from a book, which is then binarized, with the boundaries dilated and the holes filled. Three lines of the processed paragraph are shown in Figure 6(a). By checking the horizontal projection, each line is split. The first line is shown in Figure 6(b). Bounding boxes of all letters in the first line are created as shown in Figure 6(c). Segmented letters are shown in Figure 6(d).



Can functional vision be restored in blind human subjects using a microelectronic retinal prosthesis? The initial indications suggest that, yes, it is possible. However, the

(a)

Can functional vision be restored in blind human subjects

(b)

Can functional vision be restored in blind human subjects

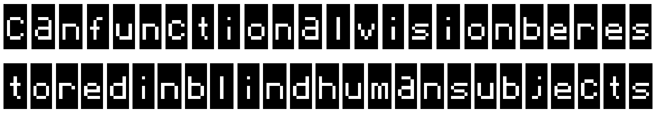
(c)

Canfunctionalvisionberest oredinblindhumansubjects

(d)

Figure 6. Results of image pre-processing. (a): Boundaries are dilated and holes are filled. (b): The split first line. (c): All letters in the first line are bounded. (d): Segmented letters are chopped sharp to the borders.

All the letters are correctly bounded, and segmented sharp to the borders, including the letters of i and j. After feature extraction and pattern recognition, all the recognized letters are represented by 5×7 pixels characters, as shown in Figure 7. These characters are then displayed in the simulated prosthetic vision, as shown in Figure 8.



canfunctionalvisionberes
toredinblindhumansubjects

Figure 7. Recognized text represented by defined characters

All the letters are recognized correctly by the BP neural network. The letters in simulated prosthetic vision are readily read. Contours of all the letters are intact and clear. These results demonstrate that this text image processing strategy is effective. In order to show its advantages in text presentation, text information processed without this strategy is presented in simulated prosthetic vision. Text images captured from the camera are converted to grayscale images. Without character recognition, all images are resized to the same resolution of 25×25. Each pixel value in the downgraded images is linear to the brightness of its corresponding phosphene. The text information is shown in Figure 9. Some contours of the text are disconnected and appear thick, making the letters more difficult to recognize. To show robustness of the strategy to grid distortion, in Figure 10 and 11, the SD of the deviations is doubled. The text with and without this strategy is shown. Text processed with this strategy is easier to recognize than the text without this strategy.

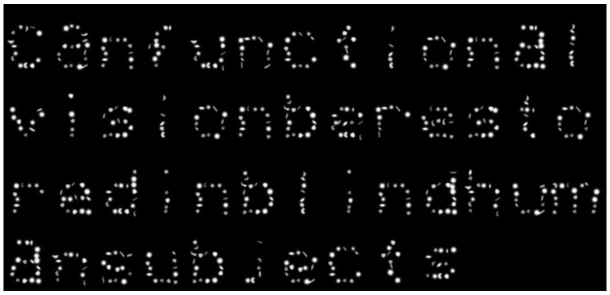


Figure 8. Text in distorted grid using this processing strategy

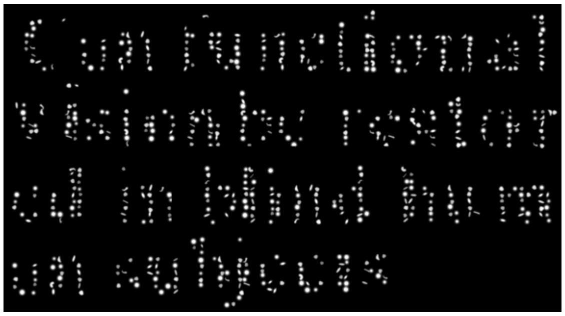


Figure 9. Text in distorted grid without using this processing strategy

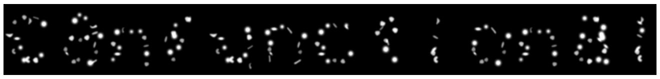


Figure 10. Text in aggravated distorted grid using this processing strategy

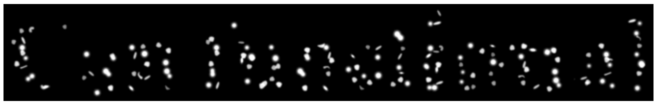


Figure 11. Text in aggravated distorted grid without this processing strategy

IV. Conclusions

We have shown the benefits of this proposed text image processing strategy. Text information in the images acquired from a camera is recognized by a feed-forward BP neural network, and then the recognized text information is represented by simplified characters of 5×7 pixels. Results of simulated prosthetic vision showed that text information was displayed correctly and read readily even using limited electrodes. These results demonstrate benefits of this proposed strategy which is aimed to provide better reading experience for blind patients using prosthetic vision.

References

[1] R. Klein, T. Peto, A. Bird et al., “The epidemiology of age-related macular degeneration,” *Am. J. Ophthalmol.*, vol. 137, no. 3, pp. 486-495, Mar, 2004.

[2] D. T. Hartong, E. L. Berson, and T. P. Dryja, “Retinitis pigmentosa,” *Lancet*, vol. 368, no. 9549, pp. 1795-1809, Nov, 2006.

[3] E. Zrenner, K. U. Bartz-Schmidt, H. Benav et al., “Subretinal electronic chips allow blind patients to read letters and combine them to words,” *Proc. R. Soc. B-Biol. Sci.*, vol. 278, no. 1711, pp. 1489-1497, 2011.

[4] J. D. Weiland, A. K. Cho, and M. S. Humayun, “Retinal prostheses: Current clinical results and future needs,” *Ophthalmology*, vol. 118, no. 11, pp. 2227-2237, Nov, 2011.

[5] M. S. Humayun, J. D. Weiland, G. Y. Fujii et al., “Visual perception in a blind subject with a chronic microelectronic retinal prosthesis,” *Vision Research*, vol. 43, no. 24, pp. 2573-2581, Nov, 2003.

[6] Y. Zhao, Y. Y. Lu, C. Q. Zhou et al., “Chinese character recognition using simulated phosphene maps,” *IOVS*, vol. 52, no. 6, pp. 3404-3412, 2011.

[7] J. Sommerhalder, "How to Restore Reading With Visual Prostheses " *Visual Prosthesis and Ophthalmic Devices*, J. Tombran-Tink, C. J. Barnstable and J. F. Rizzo, eds., pp. 15-35: Humana Press, 2007.

[8] D. Nanduri, J. D. Dorn, M. S. Humayun et al., “Percept properties of single electrode stimulation in retinal prosthesis subjects,” *IOVS*, vol. 52, no. 6, pp. 442, 2011.

[9] D. Nanduri, I. Fine, A. Horsager et al., “Frequency and amplitude modulation have different effects on the percepts elicited by retinal stimulation,” *IOVS*, vol. 53, no. 1, pp. 205-214, 2012.

[10] G. Dagnelie, D. Barnett, M. S. Humayun et al., “Paragraph text reading using a pixelized prosthetic vision simulator: Parameter dependence and task learning in free-viewing conditions,” *IOVS*, vol. 47, no. 3, pp. 1241-1250, 2006.

[11] E. Zrenner, R. Wilke, T. Zabel et al., “Psychometric analysis of visual sensations mediated by subretinal microelectrode arrays implanted into blind retinitis pigmentosa patients,” *IOVS*, vol. 48, no. 5, pp. 659,2007.

[12] G. Richard, M. Feucht, N. Bornfeld et al., “Multicenter study on acute electrical stimulation of the human retina with an epiretinal implant: Clinical results in 20 patients,” *IOVS*, vol. 46, no. 5, pp. 1143, 2005.

[13] S. C. Chen, G. J. Suaning, J. W. Morley et al., “Simulating prosthetic vision: I. Visual models of phosphenes,” *Vision Res.*, vol. 49, no. 12, pp. 1493-1506, 2009.

[14] M. S. Humayun, E. de Juan, J. D. Weiland et al., “Pattern electrical stimulation of the human retina,” *Vision Res.*, vol. 39, no. 15, pp. 2569-2576, 1999.

[15] J. D. Dorn, A. K. Ahuja, M. Arsiero et al., “The Argus II retinal prosthesis provides complex form vision for a subject blinded by retinitis pigmentosa,” *IOVS*, vol. 51, no. 5, pp. 3020, 2010.

[16] K. Cha, K. W. Horch, R. A. Normann et al., “Reading speed with a pixelized vision system,” *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.*, vol. 9, no. 5, pp. 673-677, 1992.

[17] R. W. Massof, “A systems model for low vision rehabilitation. II. Measurement of vision disabilities,” *Optom. Vis. Sci.*, vol. 75, no. 5, pp. 349-373, 1998.

[18] J. D. Weiland, and M. S. Humayun, “Visual prosthesis,” *Proceedings of the IEEE*, vol. 96, no. 7, pp. 1076-1084, Jul, 2008.

[19] D. Nanduri, M. S. Humayun, R. J. Greenberg et al., "Retinal prosthesis phosphene shape analysis," *30th IEEE EMBC*, pp. 1785-1788, 2008.

[20] J. F. Rizzo, J. Wyatt, J. Loewenstein et al., “Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials,” *IOVS*, vol. 44, no. 12, pp. 5362-5369, 2003.

[21] A. Horsager, and I. Fine, "The perceptual effects of chronic retinal stimulation," *Visual Prosthetics: Physiology, Bioengineering and Rehabilitation*, G. Dagnelie, ed., pp. 271-300, Springer, 2011.

[22] M. Bokser, “Omnidocument technologies,” *Proceedings of the IEEE*, vol. 80, no. 7, pp. 1066-1078, Jul, 1992.