Large-Capacity Image Information Reduction Based on Single-Cue Saliency Map for Retinal Prosthesis System

Yili Chen [1],Xiaokun Liang [1,2], Zhicheng Zhang [1] , Rongmao Li [1] , Nan Fu [1] , Yanchun Zhu [1] , Yaoqin Xie [1*], Huailing Zhang [2*]

1. Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences,
Research Centre for Medical Robotics and Minimally Invasive Surgical Devices.
e-mail: yq.xie@siat.ac.cn
2. *Department of Biomedical Engineering*, Guangdong Medical College.
e-mail: huailing@163.com

Abstract **- In an effort to restore visual perception in retinal diseases, an electronic retinal prosthesis with thousands of electrodes has been developed. The image processing strategies of retinal prosthesis system converts the original images from the camera to the stimulus pattern which can be interpreted by the brain. Practically, the original images are with more high resolution (256x256) than that of the stimulus pattern (such as 25x25), which causes a technical image processing challenge to do large-capacity image information reduction. In this paper, we focus on developing an efficient image processing stimulus pattern extraction algorithm by using a single cue saliency map for extracting salient objects in the image with an optimal trimming threshold. Experimental results showed that the proposed stimulus pattern extraction algorithm performs quite well for different scenes in terms of the stimulus pattern. In the algorithm performance experiment, our proposed SCSPE algorithm have almost five times of the score compared with Boyle's algorithm. Through experiments we suggested that when there are salient objects in the scene (such as the blind meet people or talking with people), the trimming threshold should be set around 0.4max, in other situations, the trimming threshold values can be set between 0.2max-0.4max to give the satisfied stimulus pattern .**

Keywords: retinal prosthesis; image processing; region of interest; saliency map; trimming threshold selection

I. INTRODUCTION

Several incurable eye diseases result in blindness for about 100,000 people every year [1]. Two most common retinal degenerative diseases are age-related macular degeneration (AMD) and retinitis pigmentosa (RP). RP is more severe and AMD is more prevalent. Electrical stimulation in human test subjects with these conditions has demonstrated the possibility of an electronic retinal prosthesis as a means of providing some degree of vision [1], [2].The Argus® II is the first retinal prosthesis approved for the treatment of patients blind from retinitis pigmentosa, receiving CE (Conformité Européenne) marking in 2011.Alphs-IMS followed closely and obtained CE marking in 2013[2],[3].

Retinal prosthesis is primarily divided into two components, one is implanted components and the other is external components. External components' main aim is doing image processing. It include a camera, an image processing unit, and bidirectional telemetry. Implanted components include the bidirectional telemetry, hermetically packaged electronics, and a multi-channel electrode array [5].

Generally, a healthy retina has over 100 million photoreceptors, however, the electrode array is currently at the scale of hundreds (typically 10x10, 25x25, 32x32 [6]). Some groups have done researches in image processing system. Wei Mao et al. simulated 32*32 pixels high resolution retinal prosthesis image processing based on FPGA [11]. Fei Guo et al. have achieved the functions of visual prosthesis image processing on LEON3 SoC (System on Chip)[8].And Armin Alaghi et al demonstrated that stochastic designs can be significantly smaller, faster, more power-efficient, and more noise-tolerant than conventional ones[9].

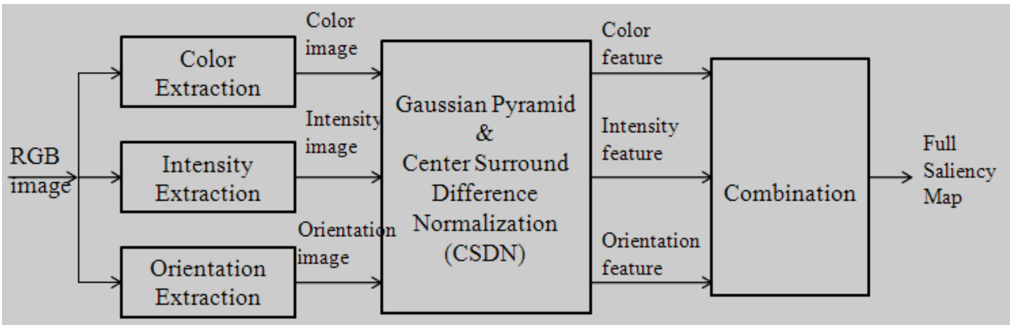
Since the image pixels should be corresponding with the stimulation electrodes, it is a real practice that the captured higher resolution images must be downscaled to lower resolution (only a few hundred pixels) so that to stimulate the corresponding electrodes. And keeping the salient objects which are defined as ROI(region of interest) area as many as possible. During this transformation process, large amounts of information may be lost[11]. As a result, it is crucial developing an efficient stimulus pattern extraction method to enhance this perception under limited resolution.

Despite lack of knowledge about processing mechanism of information from the human photoreceptor layer to the optic nerve, some research groups have already tried to evaluate the feasibility of artificial vision[12][13] Buffoni et al. has used six image processing methods, such as reduction, enhanced resolution reduction, resolution reduction and edges, binary, edges, region selection, to extract the low resolution images for the retina prosthesis system. Their research outcomes concluded that a binary method or a selected region method seems more suitable for this application. Although the image threshold method is the simplest, it leaves unwanted details that have a negative effect on the stimulus pattern image (SPI) intelligibility. On the other hand, a selected region method presents its ability to reduce the scene at different distance to a very simplistic scene representation [5].Jing Wang et al. showed that background reduction with foreground enhancement .increased response accuracy compared with methods that directly merged pixels to lower resolution [10].Boyle et al. [14], [15] emphasized on the region of interest (ROI) detection. They suggested that the ROI can be detected by the classical saliency map generation (CSMG) algorithm proposed by Itti & Koch [17].

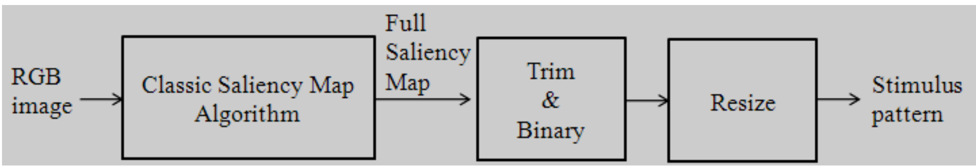
As shown in Fig 1(a), the CSMG algorithm is based on biologically motivated selective attention mechanism in human visual pathway. It uses three image cues (color, intensity, orientation) to generate a single topographical full saliency map. The research results have shown that the full saliency map image has the ability to enhance the salient objects in the original image [16]. We noted that Boyle et al. firstly applied the CSMG algorithm in the SPI extraction application for retinal prosthesis. Moreover, they have developed a trim and binary approach to get the final SPI (refer to "trim and binary" block in Fig 1(b)), where by setting the threshold as 95 percent of the maximum value of the saliency map, the full saliency map image is trimmed from their outer border until only pixels above the threshold remained. The trimmed image will be simply converted to the binary image and finally resized to the required stimulus pattern image with the resolution of 25x25. After investigating and evaluating Boyle's SPIE algorithm, we found that their method is quite complex since three-cue saliency map extraction processing need to be computed separately. For example, running time of Boyle's algorithm in MATLAB for a 256x256 input RGB image takes about 16 seconds to obtain the SPI. The high complexity prohibits its application for the real-time retinal prosthesis system. We encouraged by the fact that human eyes are more sensitivity to brightness, want to use the intensity feature alone instead of three features to generate the saliency map, and may lead to an acceptable result at a much lower computational complexity.

II. STIMULUS PATTERN EXTRACTION ALGORITHM

Our development is motivated by the capability of the CSMG algorithm for extracting the salient objects in the image, as well as the encouraging results from the research of Boyle's group. The conceptual illustration of the CSMG algorithm and the Boyle's SPI extraction algorithm is shown in Fig 1(a) and (b), respectively.



(a) Block diagram of the classical saliency map algorithm [17]



(b) Block diagram of Boyle's stimulus pattern image extraction (SPIE) Algorithm [15]

Fig 1. Conceptual Illustrations of the CSMG algorithm and the Boyle's SPI extraction algorithm.

Taking the computational complexity and stimulus pattern image (SPI) intelligibility as our main concerns, a novel single-cue stimulus pattern extraction (SCSPE) algorithm has been proposed. The block diagram of SCSPE algorithm is shown in Fig 2.

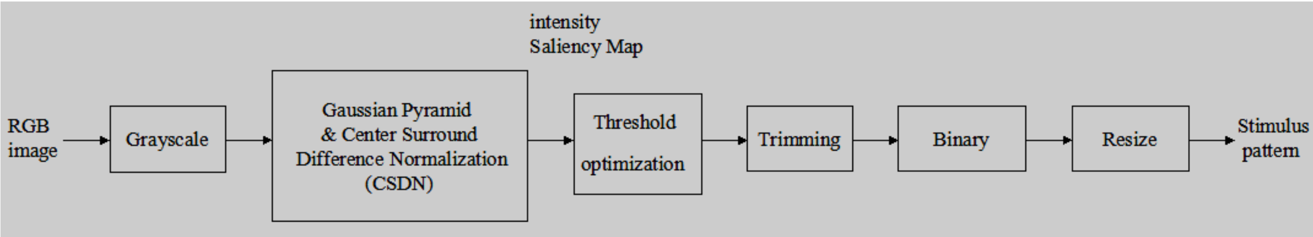


Fig 2. Block diagram of our proposed SCSPE algorithm

Where the input RGB image is obtained from the front-end camera with the resolution of 256x256.The grayscale block converts the RGB color image into the grayscale image. The Gaussian Pyramid and CSDN block calculates the intensity saliency map based on the saliency map model [16]. The optimized threshold trimming and binary block determines an optimal trimming threshold according to the image scene analysis and conducts the trimming to keep the parts of the image whose intensity saliency value is above the threshold, and also converts the trimmed image into binary image. Finally the resize block just resizes the binary image to the stimulus patterns with the resolution of 25x25 for the 625-channel retinal prosthesis system.

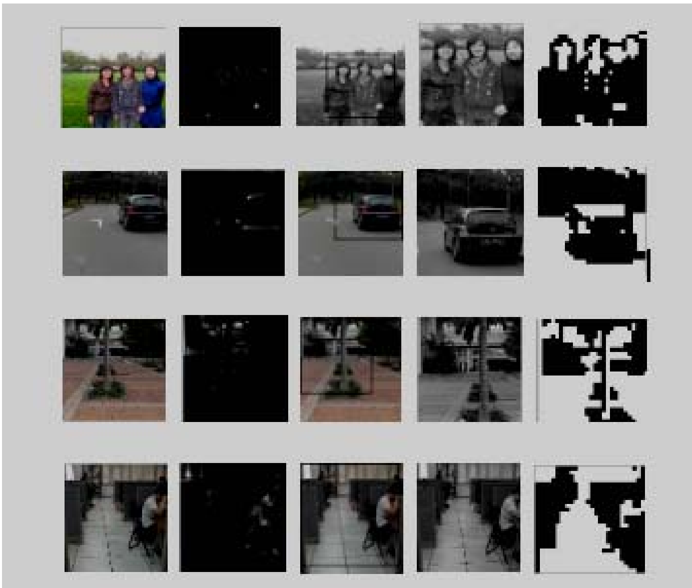


Fig 3. (From left to right) original RGB images, intensity saliency map, threshold optimization area, strimming area, binary to extracted stimulus patterns by our algorithm

Fig 3 is our proposed SCSPE algorithm results. From left to right is the original RGB images, intensity, saliency map, threshold optimization area, strimming area, binary to stimulus pattern.

In our algorithm, the intensity saliency map is generated as follows. For a grayscale image (256x256), first the six levels of Gaussian pyramid images are obtained by zooming-out and Gaussian filtering. In general, the multi-resolution Gaussian images ( ) shown in eqn. (1) is good for reducing noise and scale-invariant characteristic [19], [20].

$$I(\sigma) = G(\sigma) \bullet i \qquad (1)$$

Where G is the Gauss filter function, σ is the scale of the image, $\sigma \in \{0..6\}$; $I(\sigma)$ is the level of intensity Gaussian pyramid image.

After obtaining the six-level Gaussian pyramid images, we then choose the last five-level images to conduct the center-surround different normalization (CSDN) algorithm. In the CSDN, the five Gaussian pyramid images will first be zoom-in to the same size of the first level image in the chosen five images, and then perform the center-surround differences to get the CSD intensity feature maps, see eqn. (2). Finally add these Normalized CSD intensity feature maps to get the intensity saliency map, see eqn. (3). One example of generating the intensity saliency map is given in Fig 4.

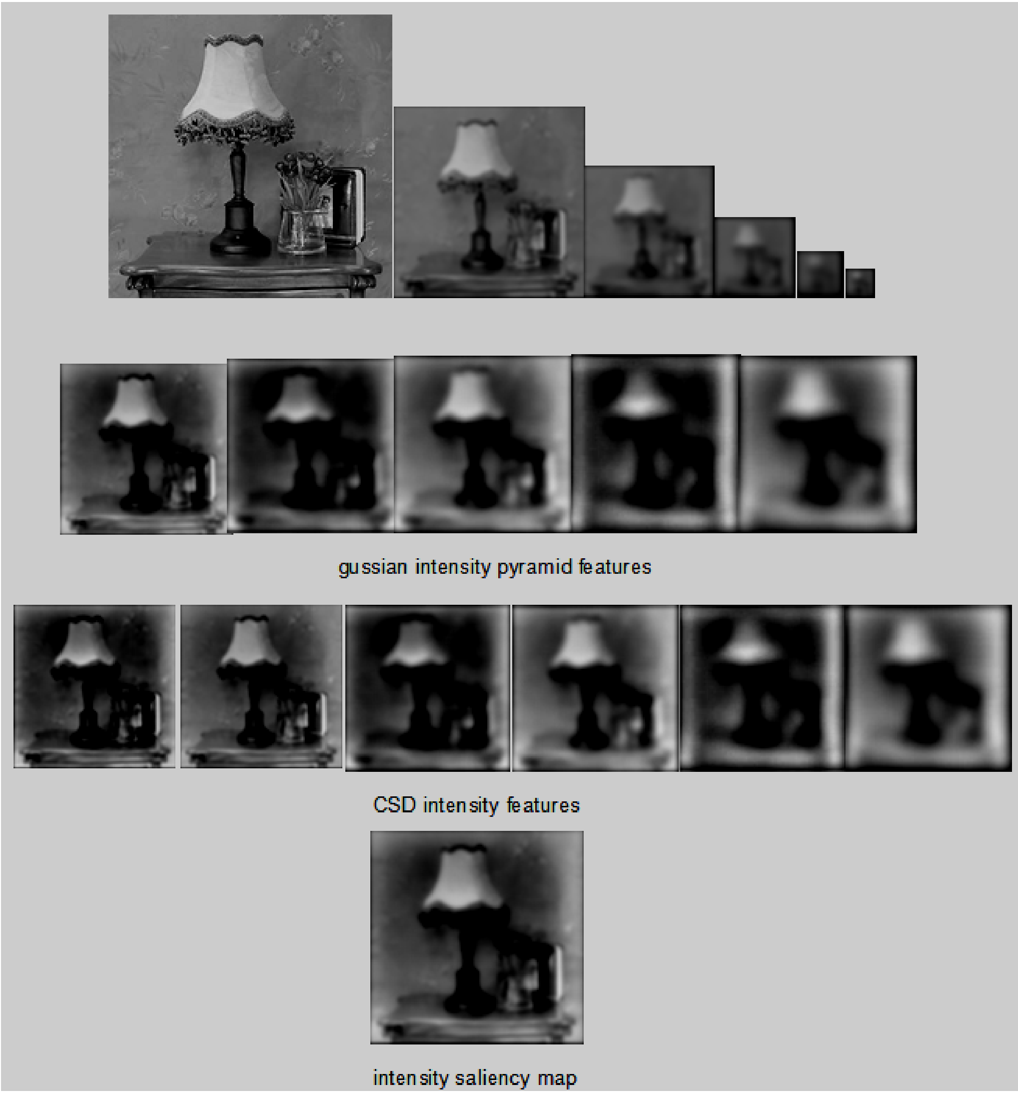


Fig 4. Generation of the intensity saliency map

The following 2 equations are proposed:

$$\bar{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{4} N(I(c,s)) \qquad (2)$$

$$I(c,s) = |I(c) - I(s)| \qquad (3)$$

in which, c is center finer scale $c \in \{2,3,4\}$, , s is the surround coarser scale, s=c+δ,$\delta \in \{3,4\}$, *I(c, s)* is the intensity difference of the center image and the surround image; $\bar{I}$ is intensity saliency map; "$\oplus$" represents an across-scale addition operation.

III. EXPERIMENTS RESULTS AND ANALYSIS

A. Input Images

In order to evaluate the performance of our single-cue stimulus pattern extraction (SCSPE) algorithm and find the optimal trimming thresholds for different scenes. We carefully selected input images of different scenes in our daily life. The images shown in Fig 5 are categorized into the indoor/outdoor, and each category includes five scenes which a blind might encounter during his/her daily life.

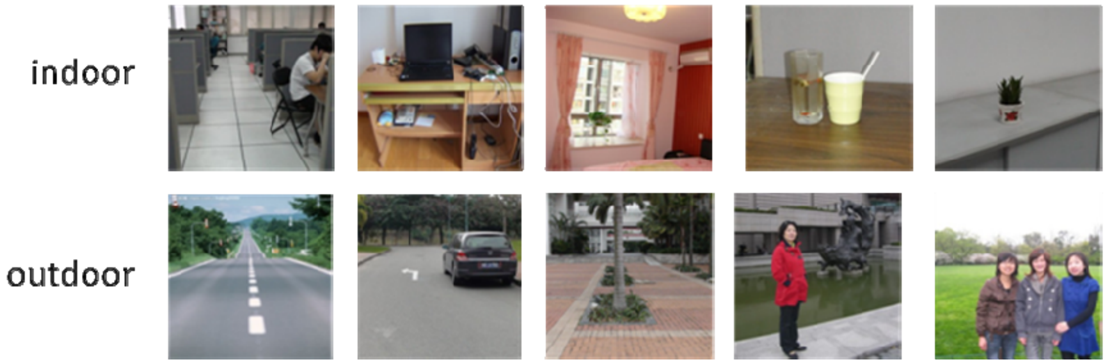


Fig 5. Input test images comprised different scenes that a blind person might encounter (256x256 RGB color images)

B. Experimental Results and Analysis

The first experiment is setup to compare the performance of our proposed SCSPE algorithm with the Boyle's algorithm. The original RGB images and the corresponding extracted stimulus patterns for four representative scenes are shown in Fig 6.

Fig 6. (From left to right) original RGB images, extracted stimulus patterns by our algorithm, extracted stimulus patterns by Boyle's algorithm.

A group of 60 normally sighted or corrected-to-normal volunteers have been invited to participate in the performance evaluation of the experimental results to further validate the research outcomes,. Subjects were presented with the original high-resolution (256x256) grayscale images and the extracted stimulus patterns by our SCSPE algorithm and Boyle's algorithm. The questionnaire has the instruction as "If you see the scene above, which version would you find most helpful?". The evaluation result is given in Fig 7.

From Fig 7, it is clear to see that, for these selected image scenes, our proposed SCSPE algorithm gives almost five times of the score compared with Boyle's algorithm, which means that the resulting stimulus patterns (low resolution images) by our algorithm can give more meaningful information and it is more suitable for the low-resolution retinal prosthesis system.

Fig 7. Performance evaluation results by 60 participates

In addition, in our SCSPE algorithm, trimming the intensity saliency map based on the calculated intensity saliency map is one important process. We noted that the selection of the trimming threshold for different image scenes becomes a problem. Experimental results showed that trimming threshold does heavily influence the resultant stimulus pattern. In order to get the better or optimal trimming threshold, we carried the following experiment.

Not only use the image as shown in Fig 5, we also randomly selected ten Boly's test pictures [15].For a given input image, the maximum gray level of the intensity saliency map is computed and it is denoted as max, then eight different trimming thresholds can be determined as 0.1max to 0.8max at the step of 0.1max.The eight different stimulus patterns using eight different trimming thresholds are calculated using proposed SCSME algorithm. In order to conduct the fair performance evaluation, we place these stimulus patterns in a random order below the original input images, which are presented in Fig 8.

In Fig 8(a), when thresholds use 0.1max to 0.6 max, the results seems good for our eyes to see the two people in the picture. But in Fig 8(b) when thresholds use 0.1max to 0.8max the results seems quite the same, may be there are no ROI area in this picture. So different pictures have different best thresholds. The same group of people has been invited to give the evaluation results as well. Viewing conditions for the experiment were not controlled. The questionnaire results are shown in Fig 9. It is noted that the results in Fig 10(a) are very interesting. Whether for indoors or outdoors scenes, the stimulus patterns using thresholds between 0.2max-0.4max received higher score. Furthermore, experimental results in FIg 9(b) showed that when there are salient objects (person\object) in the original input images, stimulus patterns with the threshold about 0.4max obtained much higher score. One of the explanations is that, for these scenes, viewers prefer to see the details of salient objects. Numerous experimental results further support above observations. From these initial research outcomes of our proposed algorithm, we suggest that when there are salient objects in the scene (such as the blind meet people or talking with people), the trimming threshold should be set around 0.4max, in other situations, the trimming threshold values can be set between 0.2max-0.4max to give the satisfied stimulus pattern .

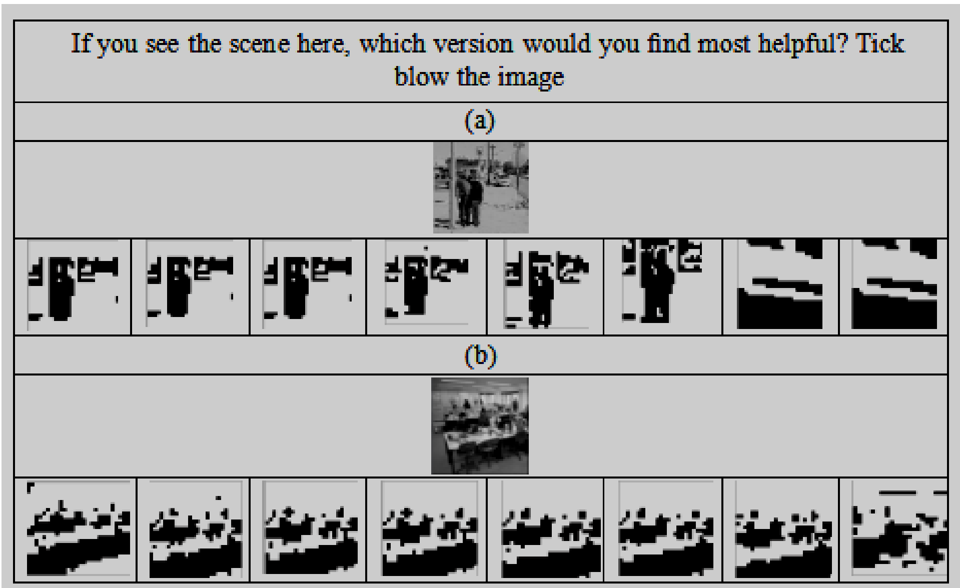
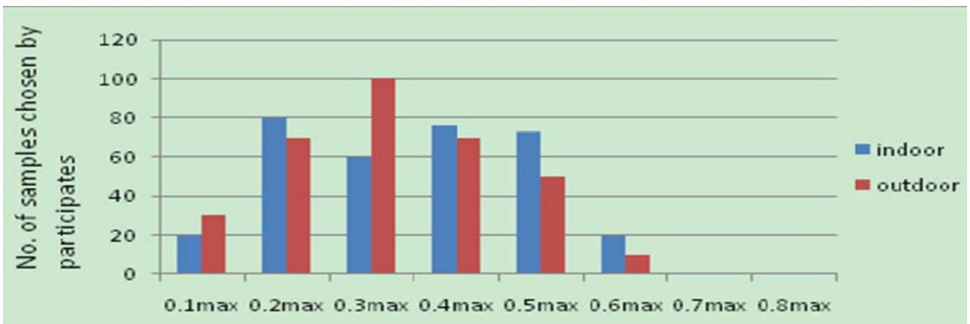
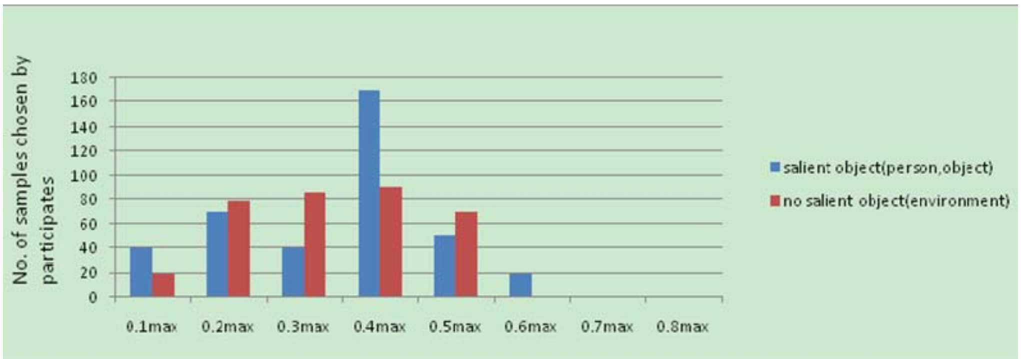


Fig 8 – Stimulus patterns by our proposed algorithm using eight different trimming thresholds



(a) Performance evaluation results of different trimming thresholds for indoor/outdoor scenes



(b) Performance evaluation results of different trimming thresholds for scenes w/o salient objects

Fig 9. Performance evaluation results by 60 participants

IV. CONCLUSIONS

A large-capacity image information reduction in real-time retina prosthesis system has been developed. This automatic extraction algorithm can be employed in a prosthesis design to highlight ROI areas that may help a visually impaired user and it has low computation complexity. Extensive experiments have been conducted to validate the performance of our proposed algorithm for different scenes. Other experiments also indicated and suggested the choice of a good threshold for different scenes. It deserves to work more in the future to develop an automatic threshold SPE algorithm or other useful algorithm for further improving the stimulus pattern extraction under different scenes for the retinal prosthesis systems.

ACKNOWLEDGEMENTS

We would like present our appreciation for the help of Boyle by providing us his test images.This work is supported in part by grants from National 863 project (No.2014AA020503,No.2015AA043203), National Natural Science Foundation of China (NSFC: 81171402, 60772120, 61471349), Guangdong Innovative Research Team Program (no. 2011S013) of China, Science and Technological Program for Dongguan's Higher Education, Science and Research, and Health Care Institutions (Grant No.2011108101001), Beijing Center for Mathematics and Information Interdisciplinary Sciences, and Comprehensive Strategic Cooperation Project of Guangdong province and Chinese academy of sciences (Grant No.2011B090300079). Yili Chen and Xiaokun Liang contributed equally to this paper.

REFERENCES

[1] M. S. Humayun, "Intraocular retinal prosthesis," American Society of Ophthalmic vol. 99, pp. 271-300, 2001.

[2] D.Weiland, Wentai Liu, and Mark S. Humayun Doheny Retina Institute, "Retinal prosthesis", in Artificial Sight, 2005.

[3] Yvonne H.-L. Luo and Lyndon da Cruz, " A review and update on the surrent status of retinal protheses (bionic eye)" ,British Medical Bulletin, 2014, 31–44, doi: 10.1093/bmb/ldu002

[4] A K Ahuja,1 J D Dorn,1 A Caspi,1 M J McMahon,1 G Dagnelie,LdaCruz,P Stanga,4 M S Humayun,5 R J Greenberg,1 Argus II Study Group, "Blind subjects implanted with the Argus Ⅱ retinal prothesis are able to improve performace in a spatial-motoe task",r J Ophthalmol 2011; 95:539 e543. doi:10.1136/bjo.2010.179622

[5] Fink W, Tarbell M, Weiland J D and Humayun, "Digital object recognition audio-assistant for the visually impaired Invest," Ophthalmol. Vis. Sci, 45: 4201,2004.

[6] Wentai Liu, Wolfgang Fink, Mark Tarbell, and Mohanasankar Sivaprakasam, "Image processing and interface for retinal visual prostheses," IEEE J. Solid-State Circuits, vol. 13, pp. 1987-1999, 2007.

[7] Wei Mao1,a, Dashun Que2,b, Huawei Chen1, Mian Yao, " research on image processing system for retinal prothesis" ,International Symposium on Computers & Informatics (ISCI 2015)

[8] Fei Guoa, Yuan Yang, Yong Gaoa, Chuan Ke Wu, " design of visual prothesis image processing system based on SOC ", International Symposium on Photonics and Optoelectronics 2014, Proc. of SPIE Vol. 9233, 92331B

[9] Armin Alaghi, Cheng Li and John P. Hayes, "Stochastic Circuits for Real- Time Image -Processing Applications",DAC 13, May 29–June 07, 2013, Austin, TX, USA

[10] Jing Wang , Yanyu Lu , Liujun Gu , Chuanqing Zhou , Xinyu Chai, " Moving object recognition under simulated prosthetic vision using background-subtraction-based image processing strategies " ,Information Sciences 277 (2014) 512–524.

[11] S. J. Park, K. H An, and M. Lee, "Saliency map model with adaptive masking based on independent component analysis," Neuro computing, vol.49, pp. 417–422, 2002.

[12] Jason Dowling. "Artificial human vision," in Artificial Sight, pp. 133-137, 2007

[13] James D. Weiland, Wolfgang Fink, Mark S. Humayun, and Wentai Liu, Wen Li, "Systems design of a high resolution retinal prosthesis," in Annu. Rev. Biomed. 2005. 7:361,401.

[14] Boyle, A. Maeder, and W. Boles, "Inherent visual information for low quality image presentation," in WDIC 2003—2003 APRS Work-shop on Digital Image Computing: Medical Applications of Image, pp. 51–56, The Australian, Recognition Society 2003.

[15] Justin R. Boyle, Anthony J. Maeder, "Region-of-interest processing for electronic visual prosthesis," Commonwealth Scientific and Industrial Research Organization, Information and Communication Technologies Center, Journal of Electronic Imaging, 17(1), 013002, 2008.

[16] Saliency map source code sourced from iLab, University of Southern California: http://ilab.usc.edu/toolkit/
[17] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Patt. Mach. Intell, vol. 20, no. 11, pp. 1254–1259, 1998.
[18] S. J. Park, K. H An, and M. Lee, "Saliency map model with adaptive masking based on independent component analysis," Neuro computing, vol.49, pp. 417–422, 2002.
[19] J.M. Ogden, E. H. Adelson, J. R. Bergen, and P. J. Burt, "Pyramid methods in image processing," RCA Engineer, vol. 30, no. 5, pp. 33–41, 1984.
[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.