A STIMULUS PATTERN EXTRACTION ALGORITHM BASED ON SALIENCY MAP FOR A 625-CHANNEL RETINAL PROSTHESIS SYSTEM

Yuexian ZOU, Yili CHEN, Yali ZHENG

Advanced Digital Signal Processing Laboratory, School of Computer & Information Engineering,
Peking University Shenzhen Graduate School
Email: zouyx@szpku.edu.cn

ABSTRACT

Electronic retinal prosthesis system has been developed to restore some vision for the blinds who lose their vision due to kind of retinal diseases. The image processing of retinal prosthesis system converts the original images from the camera to the stimulus pattern that can be properly interpreted by the brain. Practically, the original images are with much high resolution (256x256) than that of the stimulus pattern (such as 25x25), which causes a technical challenge to extract the stimulus pattern from the original image. In this paper, we focus on developing an efficient stimulus pattern extraction algorithm by using the single cue saliency map, where the salient objects in the image with an optimal trimming threshold are extracted. Experimental results show that the proposed stimulus pattern extraction algorithm performs quite well for different scenes in terms of the perception of the stimulus pattern. Some suggestions are also given on trimming threshold selection for different scenes.

Key Words*: retinal prosthesis, image processing, region of interest, saliency map, trimming threshold selection*

1. INTRODUCTION

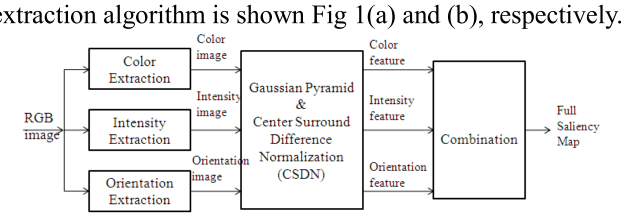
Several incurable eye diseases result in blindness for 100,000's of individuals each year [1]. Age-related macular degeneration (AMD) and retinitis pigmentosa (RP) are two most common outer retinal diseases. Experiments in human test show that the subjects who has the disease of AMD or RP can observe light spot or line by implanting the electronic retinal prosthesis system, which demonstrated the feasibility of the electronic retinal prosthesis system to provide some vision for the blinds [1], [2].

The retinal prosthesis system primarily consists of the implanted and external subsystems. External subsystem composes a camera, an image processing unit, and bidirectional telemetry. The implanted components include the bidirectional telemetry, the microstimulator chip and a multi-channel electrode array [3].

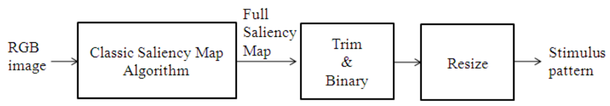
The image processing unit of the retinal prosthesis system is one of the key parts of the prosthesis. Generally, a healthy retina has over 100 million photoreceptors, however, the electrode array is currently at the scale of hundreds (typically 10x10, 25x25, 32x32 [4]). Since the image pixels should be corresponding with the stimulation electrodes, it is a common practice that the captured higher resolution images must be downscaled to lower resolution (only a few hundred pixels) but keeping the salient objects as many as possible. During this transformation process, large amounts of information will be lost. As the result, it is crucial to enhance this perception under limited resolution by developing an efficient stimulus pattern extraction method.

Despite the lack of knowledge about processing of information from the human photoreceptor layer to the optic nerve, some research groups have already tried to evaluate the feasibility of artificial vision from an image processing point of view. 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 [3].

Boyle et al. [6], [8] 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 [10]. The conceptual illustration of the CSMG algorithm and the Boyle's SPI extraction algorithm is shown Fig 1(a) and (b), respectively.



(a) Block diagram of the classical saliency map algorithm [10]



(b) Block diagram of Boyle's stimulus pattern image extraction (SPIE) Algorithm [8]

Fig 1– Conceptual Illustrations.

As shown in Fig 1(a), the CSMG algorithm is based on biologically motivated selective attention mechanism in human visual pathway. Three image cues (color, intensity, orientation) have been used to generate a single topographical full saliency map. The research has shown that the full saliency map image has the ability to enhance the salient objects in the original image [9]. 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 are 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 its complexity is quite high 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 give the SPI. The high complexity prohibits its application for the real-time retinal prosthesis system. Encouraged by the fact that human eyes are more sensitivity to brightness, using the intensity feature alone instead of three features to generate the saliency map may lead to an acceptable result at a much lower computational complexity.

2. 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. 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

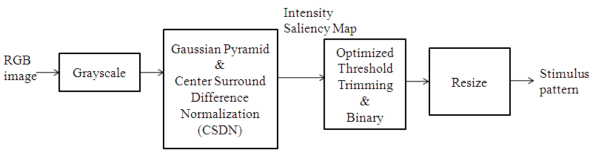


Fig 2 – Block diagram of our proposed SCSPE algorithm

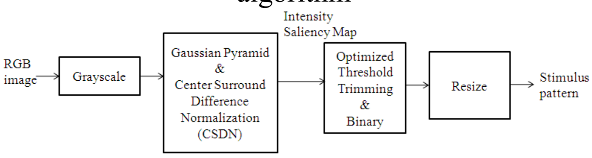


Fig 2, the input RGB image is from the front-end camera with the resolution of 256x256.The grayscale block converts the RGB color image into the grayscale image. The Gaussian Pyramid and CSDN block calculates the intensity saliency map based on the saliency map model [9]. The optimized threshold trimming and binary block determines a optimal trimming threshold according to the image scene analysis and conduct 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.

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 ($I(\sigma)$) shown in eqn. (1) can be used to provide good performance for reserving the scale-invariant characteristic and reducing noise [12], [13].

$$I(\sigma) = G(\sigma) \bullet i \tag{1}$$

In (1), G is the Gaussian filter function, σ is the scale of the image, $\sigma \in \{0..6\}$; $I(\sigma)$ is the σ level of intensity Gaussian pyramid image.

After obtaining the six-level Gaussian pyramid images, then we 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, as shown in eqn. (2). Finally the summation of these Normalized CSD intensity feature maps generate the intensity saliency map, as denoted in eqn. (3). One example of generating the intensity saliency map is illustrated in Fig 3 ([6], [11]).

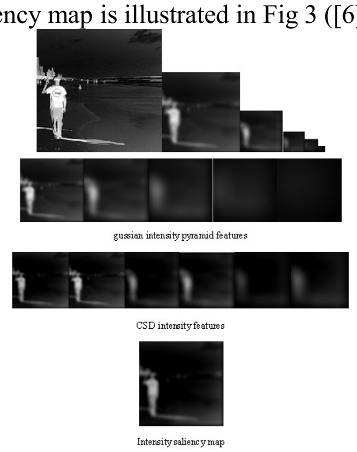


Fig 3 – The generation of the intensity saliency map

$$I(c,s) = | I(c) - I(s) | \tag{2}$$

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

In which, c is the finer scale of the center ($c \in \{2,3,4\}$), s is the surround coarser scale, $s = c + \delta, \delta \in \{3,4\}$, 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.

For comparison purpose, the experimental results of the saliency map generated from Boyle's algorithm and the intensity saliency map are shown in Fig 4 (a) and Fig 4 (b), respectively. The original RGB color image (256x256) is also presented in Fig 4 (c).

Fig 4 – (a) Boyle's saliency map image; (b) The intensity saliency map image; (c) The original RGB color image

From Fig 4, it can be seen that there is no big difference between Boyle's saliency map generating by three cues and the intensity saliency map generating by single-cue (intensity). To solidify our idea, a number of experiments have been carried out for computing the saliency map using different image cues. It is encouraged to see that the simulation results validate our considerations that for majority images in this low-resolution application, it is desired to see that the stimulus pattern image extraction using single-cue saliency map has much lower computation complexity than Boyle's algorithm and it is more preferable for the real-time retinal prosthesis system.

3. EXPERIMENTS RESULTS AND ANALYSIS

3.1 Input Images

In order to evaluate the performance of the proposed single-cue stimulus pattern extraction (SCSPE) algorithm and find the optimal trimming thresholds for different scenes, input images of different scenes have been carefully selected. 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.

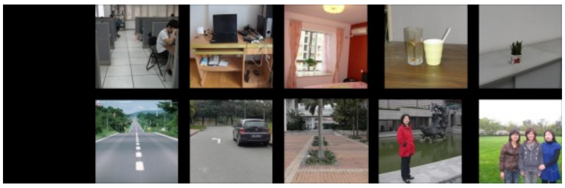


Fig 5 – Input test images comprised different scenes that a blind person might encounter (256x256 RGB color images)

3.2 Experimental Results and Analysis

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

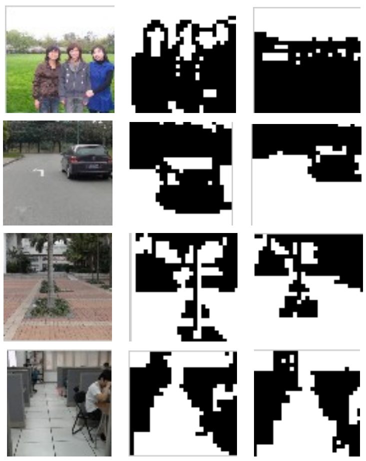


Fig 6 – (From left to right) original RGB images, extracted stimulus patterns by our algorithm, extracted stimulus patterns by Boyle's algorithm.

To further validate the research outcomes, a group of 60 normally sighted or corrected-to-normal volunteers have been invited to participate in the performance evaluation of the experimental results. 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.

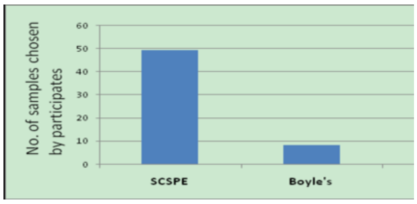


Fig 7 – Performance evaluation results by 60 participates

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 may give more meaningful information and it is more suitable for the low-resolution retinal prosthesis system.

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 present in Fig 5, we also random select ten Boly's test pictures [8].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 0.1*max* to 0.8*max* at the step of 0.1*max*.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. The same group of people has been invited to give the evaluation results as well. The viewing conditions for the experiments were not under control. The questionnaire results are shown in Fig 9. It is noted that the results in Fig 9(a) are very interesting. Whether for indoors or outdoors scenes, the stimulus patterns using thresholds between 0.2*max*-0.4*max* 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 0.4*max* 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 may 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 0.4*max*, in other situations, the trimming threshold values can be set between 0.2*max*-0.4*max* to give the satisfied stimulus pattern .

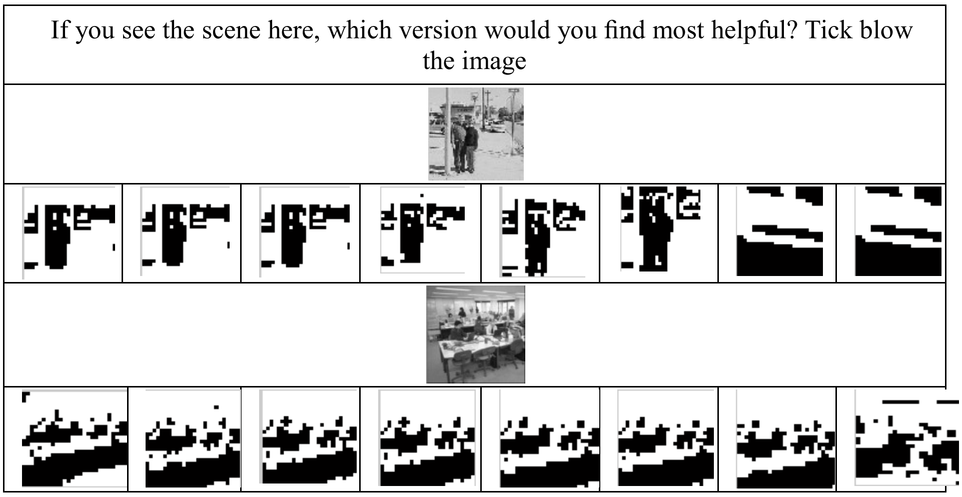
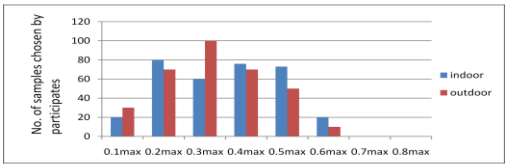
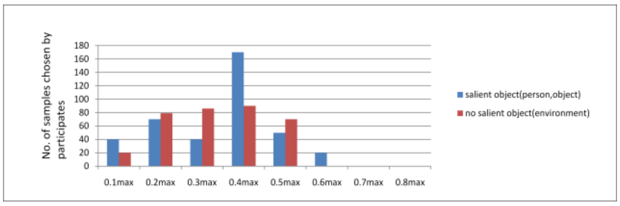


Fig 8 – Stimulus patterns by our proposed algorithm using eight different trimming thresholds



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



(b) Performance evaluation results of different trimming thresholds for scenes with salient/no salient objects

Fig 9 – Performance evaluation results by 60 participants

4. CONCLUSIONS

A stimulus pattern extraction algorithm for a real-time retina prosthesis system has been developed. This stimulus pattern automatic extraction algorithm can be employed in a prosthesis design to highlight areas that may help a visually impaired user. 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 the good thresholds for different scenes. It deserves to work more in the future to develop an automatic threshold SPE algorithm for further improving the stimulus pattern extraction under different scenes for the retinal prosthesis systems automatically.

ACKNOWLEDGMENTS

We would like present our appreciation for the help of Boyle by providing us his test images. This work is supported by National High Technology Research and Development Program of China (863 Program No.2007-AA11Z224), and Shenzhen Nanshan Bureau of Science Technology and Information Program.

REFERENCFES

[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] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] Saliency map source code sourced from iLab, University of Southern California: http://ilab.usc.edu/toolkit/

[10] 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.

[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] 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.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004.