G Model
ARTMED-1560; No. of Pages 15
ARTICLE IN PRESS
Artificial Intelligence in Medicine xxx (2017) xxx–xxx
Contents lists available at ScienceDirect
Artificial Intelligence in Medicine
journal homepage: www.elsevier.com/locate/aiim

Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision

Heng Li [a], Xiaofan Su [a], Jing Wang [b], Han Kan [a], Tingting Han [a], Yajie Zeng [a], Xinyu Chai [a,*]

[a] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[b] College of Information Technology, Shanghai Ocean University, Shanghai 201306, China

ARTICLE INFO

Article history:
Received 9 October 2016
Received in revised form 3 November 2017
Accepted 7 November 2017

Keywords:
Visual prosthesis
Simulated prosthetic vision
Saliency segmentation
Image processing strategy
Objects recognition

ABSTRACT

Background and objective: Current retinal prostheses can only generate low-resolution visual percepts constituted of limited phosphenes which are elicited by an electrode array and with uncontrollable color and restricted grayscale. Under this visual perception, prosthetic recipients can just complete some simple visual tasks, but more complex tasks like face identification/object recognition are extremely difficult. Therefore, it is necessary to investigate and apply image processing strategies for optimizing the visual perception of the recipients. This study focuses on recognition of the object of interest employing simulated prosthetic vision.
Method: We used a saliency segmentation method based on a biologically plausible graph-based visual saliency model and a grabCut-based self-adaptive-iterative optimization framework to automatically extract foreground objects. Based on this, two image processing strategies, Addition of Separate Pixelization and Background Pixel Shrink, were further utilized to enhance the extracted foreground objects.
Results: i) The results showed by verification of psychophysical experiments that under simulated prosthetic vision, both strategies had marked advantages over Direct Pixelization in terms of recognition accuracy and efficiency. ii) We also found that recognition performance under two strategies was tied to the segmentation results and was affected positively by the paired-interrelated objects in the scene.
Conclusion: The use of the saliency segmentation method and image processing strategies can automatically extract and enhance foreground objects, and significantly improve object recognition performance towards recipients implanted a high-density implant.

1. Introduction

Retinal prostheses have been a viable way to restore partial vision in blind patients undergoing retinal diseases, such as retinitis pigmentosa (RP) and potentially dry age-related macular degeneration (AMD) [1]. For these blinding diseases, the photoreceptors suffer from degeneration while the inner retinal cells (e.g. ganglion cells and bipolar cells) remain largely intact in terms of their morphology and transmission of visual information [2]. By using a surgically implanted electrode array to electrically stimulate these surviving neurons in the retina, the “phosphenes” (dots or spots of light), the basic units or elements of prosthetic vision perception, can be elicited reliably and reproducibly [3–6].

According to the anatomical location where the electrode array is implanted, there are currently three types of retinal prostheses that have been developed: epiretinal, subretinal and suprachoroidal prostheses (read reviews [7–9] on the details of their design and working principles, and on the benefits and challenges in their engineering and surgery). Amongst them, there are two commercially available models: the Argus II epiretinal prosthesis developed by Second Sight Medical Products (obtained CE marking in Europe in March 2011 and FDA approval in the USA in February 2013) and the Alpha-IMS subretinal prosthesis from Retinal Implant AG (received CE marking in Europe in July 2013) [1]. The number of implantable electrodes has increased from 16 (Argus I and the first generation of Retinal Implant AG devices) to 60 (Argus II) and 1500 diodes/electrodes (Alpha-IMS). Up to this date, the number of patients implanted above devices has been more than 250 cases [10,11]. A majority of subjects perform significantly better with the system on than off in functionally visual tests, including target localization, motion detection, orientation, and mobility.

* Corresponding author at: 800 Dongchuan RD, Minghang District, Shanghai, 200240, China.
E-mail address: xychai@sjtu.edu.cn (X. Chai).

https://doi.org/10.1016/j.artmed.2017.11.001
0933-3657/© 2017 Published by Elsevier B.V.

Although there has been promising progress in retinal prostheses, due to limitations imposed by current technical, physical and biological challenges [12], the visual acuity elicited by current electrode arrays is still far worse than in normal vision. Recent clinical trials reported that the best corrected visual acuity measurement was 20/1262 in patients implanted with Argus II [13] and 20/546 in patients implanted with Alpha-IMS [14], which are both lower than the threshold for legal blindness (20/200) [2]. Although the implanted patients were able to perform visual tasks such as reading letters or words, discriminating high-contrast geometric shapes, and identifying direction of motion [5,14–19], they struggled with more complicated tasks like face or object recognition, reading texts, and independent locomotion in unfamiliar complex surroundings, which suggests a need to increase the number of electrodes in future designs [20–23]. At present, there been various high-density implants being designed or tested in clinical trials awaiting further development and approval for chronic study [24–31]. For example, recently, Palanker's group at Stanford University (Palo Alto, CA, USA), for restoration of sight with high visual acuity, is close to implementing an implantable photovoltaic retinal device that can produce well-confined electric fields, with low crosstalk between compact neighboring pixels by designing distributed return localized to each stimulating electrode [31]. A rat trial conducted by the group suggested that the high-density design was able to provide an inspiring visual acuity [24,32]. In addition to the ongoing electrode approaches, optogenetic retinal prosthesis may be a more promising method for restoration of high visual acuity. This technology is based on the photosensitization of neurons using light-gated ion channels and pumps [33], which have been successfully expressed in RGCs [34] and bipolar cells [35]. Therefore, it is possible to target a single one of above cells. Besides, a study by Nirenberg and Pandarinath [36] suggested that driving the optogenetic transducers with the retina's neural code was able to dramatically increase prosthetic performance, well beyond what could be achieved just by increasing resolution. It implies that the combination of the neural code and high-resolution optogenetic stimulators is expected to bring prosthetic capabilities up to the level of normal or near-normal image representation.

In addition to the challenges in visual acuity, significant vision features provided by current retinal implants are mostly lost. For instance, the nature and stability of the electrically-elicited color percepts is still unclear, so that it is very difficult to generate controlled color vision; and perceptible brightness are generally confined within ten levels [4,5,37]. The result is that the low-resolution visual perception elicited by current retinal prostheses loses some necessary image saliency, like color, intensity, texture and edge, which can help humans detect and recognize rapidly the objects of interest in their visual scenes according to the selective visual attention mechanism [38]. These different clinical trials have demonstrated the fact that current prosthetic vision is low resolution with poor visual perception [39].

For most retinal implants, an external camera and a video processor are important components of them, and allows image processing to potentially improve understanding of visual perception for recipients [40]. The IMI (Intelligent Medical Implant) system integrated a learning algorithm that simplified tuning certain parameters by the user via an external processing algorithm [41]. The goal is the capacity to fine tune the electrical stimulation through a so-called "learning retinal encoder" proposed by Eckmiller [42,43]. The Argus II study group reported that the recipients' visual acuity was able to be improved to 20/200 with 16× magnification, and that the subject used a magnification of 4× and was able to read accurately short words from a notebook at a distance of 30 cm [44]. The group was also reported that 5 patients implanted with the Argus II System were able to find accurately the human face by a face detection processing [45]. Recently, Barnes et al. [46] reported a vision function testing of applying image filtering for a suprachoroidal retinal prosthesis (with 20 simulating electrodes). And results showed that the use of image filtering strategy could improve significantly recipients' performance in finishing with some simple visual tasks. In addition, there were many simulated prosthetic vision (SPV) studies that developed and assessed various image processing strategies directing at some difficult visual tasks, such as navigation, object recognition, and face identification [47–52]. Among these tasks, object detection or recognition is one of the most basic visual tasks in human life in which visual attention plays a very significant role. Human attention is often directed toward visual stimuli that have notable features, such as color, intensity, contrast, and orientation [53], which can help people locate and predictably fixate objects amidst complex surroundings rapidly and correctly. However, these vision features are largely weakened under the discrete low-resolution prosthetic vision [37]. In other words, prosthesis recipients' voluntary ability of the selective attention has greatly degenerated, making object detection and recognition very difficult for them.

Given the effect of selective attention on object detection and recognition, some researchers investigated the application of various attention models on simulated prosthetic vision (SPV) processing. Li et al. [54] proposed a computational pixelization model that assigned higher resolution to areas of an image with prominent features, including contrast, edge, orientation, and symmetry. The model proved effective for detecting areas that humans are interested in. Boyle et al. [55] applied region-of-interest (ROI) processing to binary images and found that the method, when used in a zoom application, could improve scene recognition performance. Van Rheede et al. [56] tested three fixed-gaze image representation methods – full-field, ROI, and fisheye – on various tasks and concluded that different tasks required corresponding image representation methods. Stacey et al. [57] developed a salient information processing system for obstacle avoidance under SPV, and results demonstrated that the system was able to identify effectively obstacles. Parikh et al. [58] utilized a saliency-based cueing algorithm for navigating tasks, and results showed that using saliency models would aid prosthesis wearers in detecting important objects. Recently, Wang et al. [59] proposed two image processing strategies using a saliency segmentation method based on the Itti saliency model and grabCut for object recognition under SPV, and results indicated that the proposed image processing strategies could improve recognition accuracy for a single object. The above findings have suggested that saliency models or feature integration based on visual attention mechanisms can be used to emphasize the regions with significant features in an image or a scene, so as to provide prosthesis recipients with rich visual perception.

This paper focuses on recognition of the object of interest (OOI) under low-resolution prosthetic vision. Considering the loss of most vision features and information in low-resolution prosthetic vision, we used a saliency segmentation method for automatically extracting the OOIs. First, a bottom-up graph-based visual saliency (GBVS) model was utilized on images with one or two objects to detect the saliency map, which was combined with percentile extraction and margin localization to automatically label a foreground region. It was used to take the place of the user-specified foreground to serve as the hard constraints (prior knowledge) of grabCut. Finally, a self-adaptive-iterative optimization framework based on grabCut (SAIF-grabCut) was developed to extract the OOI as completely as possible.

The objective of the present study was to enhance the perception of OOIs so as to improve object recognition performance in low-resolution prosthetic vision. Based on the above method, thus, we further presented two image enhancement strategies to highlight the OOI. To verify the proposed image processing strate-