ARTICLE IN PRESS

Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Moving object recognition under simulated prosthetic vision using background-subtraction-based image processing strategies

Jing Wang [a], Yanyu Lu [b], Liujun Gu [a], Chuanqing Zhou [a], Xinyu Chai [a,*]

[a] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[b] School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 2 August 2012
Received in revised form 14 January 2014
Accepted 25 February 2014
Available online xxxx

Keywords:
Visual prosthesis
Simulated prosthetic vision
Background subtraction
Image processing strategy
Moving objects recognition

ABSTRACT

A visual prosthesis that applies electrical stimulation to different parts of the visual pathway has been proposed as a viable approach to restore functional vision. However, the created percept is currently limited due to the low-resolution images elicited from a limited number of stimulating electrodes. Thus, methods to optimize the visual percepts providing useful visual information are being considered. We used two image-processing strategies based on a novel background subtraction technique to optimize the content of dynamic scenes of daily life. Psychophysical results showed that background reduction, or background reduction with foreground enhancement, increased response accuracy compared with methods that directly merged pixels to lower resolution. By adding more gray scale information, a background reduction/foreground enhancement strategy resulted in the best performance and highest recognition accuracy. Further development of image-processing modules for a visual prosthesis based on these results will assist implant recipients to avoid dangerous situations and attain independent mobility in daily life.

1. Introduction

There are no effective clinical treatments to restore vision for some retinal diseases such as age-related macular degeneration and retinitis pigmentosa. Implanting a visual prosthesis has been proposed as a viable approach to restore partial vision to blind patients suffering from these diseases. The perception of spots of light, called phosphenes, are elicited by electrically stimulating different parts of the visual pathway (retina, optic nerve, or cortex) [42,47]. Over recent decades, several research groups have developed different types of visual prosthetic devices and successfully implanted them in blind patients [7,27,40,48]. Although visual prostheses have gained significant development and continue to achieve encouraging improvement, some engineering challenges, such as electrode fabrication, power consumption, and long-term viability [23,42], remain to be overcome before microelectronic high density electrode implants can be realized. Consequently, visual perception generated by a limited number of stimulation contacts is still poor relative to normal vision. Methods to optimize the image quality presented by such a limited number of phosphene dots to maximize visual percepts are currently being considered.

* Corresponding author. Tel.: +86 021 34204077; fax: +86 021 34204078.
E-mail address: xychai@sjtu.edu.cn (X. Chai).

http://dx.doi.org/10.1016/j.ins.2014.02.136

Simulation of prosthetic vision offers an alternative way of adjusting implant designs and estimate the minimal information requirements for the prosthetic wearer. Currently used models of prosthetic vision simulate phosphene shape, intensity, size and regularity to simplify the possible range of percepts by normally sighted volunteers, and therefore, represent a best case scenario for a blind patient under these conditions [15]. In this way, many research groups have effectively estimated the performance capacity of a visual prosthesis for such things as facial recognition [36] and eccentric reading [33]. In order to provide prosthesis wearers with useful artificial vision, some researchers have incorporated and estimated several image processing strategies applicable to some basic visual functions in addition to the essential processing of image downsampling and pixelization. Boyle et al. [5] studied the effect of image processing factors on object and face recognition, which included spatial resolution, grayscale, contrast, edge detection, distance and importance mapping, in order to enhance the information content presented by limited numbers of electrodes. Dowling et al. [22] reviewed aspects related to blind mobility and discussed a variety of image processing techniques suitable for a visual prosthesis, and presented a mobility display framework. Based on previous work, Boyle et al. [6] applied several variations to region-of-interest (ROI) processing and edge detection for scene and face recognition. Results showed that presenting a digital zoom of a salient region within the generated ROI was preferable to presenting an entire ROI-processed image. Zhao et al. [44] used two kinds of image processing strategies (adaptive threshold-binarization and edge extraction) to investigate the effect of pixel shape and resolution in a recognition task of common objects or scenes. Results demonstrated that image mode had a significant impact on recognition accuracy near the threshold of recognition. van Rheede and colleagues [39] have implemented three techniques (Full-field representation, ROI and Fisheye) for optimizing the information content of a simulated prosthetic image in order to evaluate visual acuity, object recognition and manipulation, and navigation within the environment. Their experimental results suggested that changing the conditions of image presentation proved advantageous for different visual functional tasks.

Static object and scene recognition have been extensively studied, however, the capacity to perceive moving objects under circumstances such as danger avoidance and independent mobility, which is an important aspect of vision [17], have rarely been investigated with simulated prosthetic vision. McCarthy and Barnes [32] proposed a time-to-contact map based on depth image by focusing on free-moving incoming object perception. Their quantitative (sequence analysis) and qualitative (image presentations) simulations demonstrated the effectiveness of the proposed representation method for emphasizing objects posing an imminent threat of collision (via increased phosphene brightness). To facilitate the perception of moving objects in a general dynamic scene by a prosthesis wearer, we used two image processing strategies based on a universal background subtraction technique with an adaptive post-processing method. The effect and processing speed of these strategies were evaluated by a series of psychophysical experiments on the perception of a dynamic scene. We demonstrate the usefulness of background-subtraction-based image processing strategies for recognition of moving objects that simulate the experience of a blind patient implanted with a prosthesis.

2. Material and methods

2.1. Materials

Ten different dynamic scenarios (five indoors, five outdoors) were filmed and included images of people during daily activities or different kinds of moving vehicles, i.e. situations that visually impaired people might expect to routinely encounter. Each video sequence was $\sim$10 s in duration, in which the vertical viewing angle of every object (human or vehicle) was $\sim$ equal. A $20^\circ \times 20^\circ$ field of view was adopted based on the available visual field of a retinal prosthesis prototype [46]. Indoor scenes showed people presented in a vertical view that covered a visual angle of 16°–18°, whereas moving objects in an outdoor scene were presented within a 3°–5° visual angle so that all the observed moving objects remained within the entire 20° field of view. The playing speed of all scenes was set to 20 frames per second (examples of screenshots from the original video materials are shown in Fig. 1).

2.2. Image processing strategies

RGB color video images were resized to a 480×480 resolution, converted to grayscale, and then adjusted for uniform illumination. In general, the image processing stage of the visual prosthesis adjusts the image resolution by combining a set number of pixels into a single output pixel for stimulating the tissue interface array [18] and is called Merging Pixels to Low Resolution (MPLR). The small electrode number leads to a huge loss of information when presenting a real-life scene. Increasing the contrast between a moving object(s) (i.e. the foreground) and the relatively static or slow moving parts of the scene (i.e. the background) can enhance the perception of the main information. Therefore, moving objects in a dynamic scene need to be automatically detected and precisely separated from the surrounding information, i.e. a series of foreground segmentation images must be generated first from the original video. A common motion detection technique for distinguishing the foreground from the background in computer vision is background subtraction (BS). We applied a novel BS algorithm for foreground extraction and developed a post-processing method for optimizing the segmentation. Two image-processing strategies based on BS segmentation were used to increase the contrast between the moving foreground and background (Fig. 2), and then compared with direct MPLR processing.

Please cite this article in press as: J. Wang et al., Moving object recognition under simulated prosthetic vision using background-subtraction-based image processing strategies, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.02.136