Information Sciences 180 (2010) 2915–2924

Contents lists available at ScienceDirect

Information Sciences

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

# Image processing based recognition of images with a limited number of pixels using simulated prosthetic vision

Ying Zhao [a,1], Yanyu Lu [a,1], Yukun Tian [a], Liming Li [a], Qiushi Ren [b], Xinyu Chai [a,*]

[a] Department of Biomedical Engineering, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 200240 Shanghai, China
[b] Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100871, China

ARTICLE INFO

*Article history:*
Received 18 May 2009
Received in revised form 16 April 2010
Accepted 20 April 2010

*Keywords:*
Pixelated image
Visual prostheses
Low resolution
Recognition
Simulated prosthetic vision

ABSTRACT

Visual prostheses based on micro-electronic technologies and biomedical engineering have been demonstrated to restore vision to blind individuals. It is necessary to determine the minimum requirements to achieve useful artificial vision for image recognition. To find the primary factors in common object and scene images recognition and optimize the recognition accuracy on low resolution images using image processing strategies, we investigate the effects of two kinds of image processing methods, two common shapes of pixels (square and circular) and six resolutions ($8 \times 8$, $16 \times 16$, $24 \times 24$, $32 \times 32$, $48 \times 48$ and $64 \times 64$). The results showed that the mean recognition accuracy increased with the number of pixels. The recognition threshold for objects was within the interval of $16 \times 16$ to $24 \times 24$ pixels. For simple scenes, it was between $32 \times 32$ and $48 \times 48$ pixels. Near the threshold of recognition, different image modes had great impact on recognition accuracy. The images with "threshold pixel number and binarization-circular points" produced the best recognition results.



## 1. Introduction

For many diseases that have no effective treatment such as age related macular degeneration (AMD) and retinitis pigmentosa (RP), visual prostheses offer another feasible method to restore partial functional vision for blind individuals. Prostheses electrically stimulate the different parts (retina, optic nerve or cortex) of the visual pathway and elicit the dots of light perception called phosphenes [9]. Although researchers have tried to build intelligent model to mimic the brain [2,1], how the implant recipient interprets the information formed by phosphenes is still a problem of visual comprehension. Due to the limitations of the existing techniques, such as electrode size, power consumption, heat dissipation, biocompatibility etc., it will be difficult to increase the pixel number of the actual visual prosthesis. Therefore, the current goal of visual prostheses is not to regenerate fully detailed vision, but to provide visual perception that is useful for blind individuals to perform common daily tasks such as text reading, facial recognition and unfamiliar environment navigation [28]. It is necessary to determine the minimal requirements for a visual prosthesis wearer to perform such common daily tasks.

Based on the purpose and physiological results [4,37,20,31,12,38], some investigators have studied simulated prosthetic vision in normal sighted subjects using limited and different low resolution pixelated images. Inspiring results were reported

* Corresponding author. Tel.: +86 021 34204077; fax: +86 021 34204078.
*E-mail address:* xychai@sjtu.edu.cn (X. Chai).
[1] These two authors equally contributed to this work.

0020-0255/$ - see front matter 
doi:10.1016/j.ins.2010.04.021

on Latin word recognition [24,25,34], Chinese character recognition [8], reading speed [5–7], text or paragraph reading [35,15,10], maze navigation [6], object recognition [19], hand-eye coordination [19], way finding [11], etc. However, it is important to confirm the minimum requirements for visual prostheses through different experiments. Boyle et al. [3] investigated the quantity and type of information that are needed to recognize or perceive a scene. Thompson et al. [36] researched facial recognition using pixelated vision and focused on the impact of different factors (grid size, dot size, gap width, dot drop-out rate, and gray level resolution) on recognition accuracy. In cortical visual prostheses, Cha et al. [5] used simulated arrays varying from 100 ($10 \times 10$ array) to 1024 ($32 \times 32$ array), presented by small dots in a video display, and they found that a visual sense composed of about 625 ($25 \times 25$ array) discrete phosphenes within a visual field of 1.7° provides 20/30 visual acuity. Moreover, the same number of dots with a field of view of 30° provided useful visually guided mobility. However, conclusions have not yet been drawn regarding what kinds of image mode could optimize the recognition accuracy in low resolution based on image processing technologies. Furthermore, recognizing only a few simple objects is not enough in developing visual prostheses, as the recognition of daily scenes is also important to the blind in their daily life.

The purpose of this study is to investigate the primary factors in common object and scene image recognition as well as to optimize the recognition accuracy at low resolution using image processing technologies. Therefore, two image processing methods ('threshold-binarization' images and 'edge' images) and two kinds of pixel shapes (square and circular pixel) were adopted in different sized image arrays ($8 \times 8$, $16 \times 16$, $24 \times 24$, $32 \times 32$, $48 \times 48$ and $64 \times 64$). The recognition accuracy of these objects and scenes with different pixel numbers and processing methods was investigated by using a simulated prosthetic vision system in normal sighted individuals.

## 2. Materials and methods

### 2.1. Materials

Twenty object and five scene images were chosen from the databases which were familiar to everyone. The original images were captured using the mini CMOS image sensor (OV6650FS, Omnivision Inc.) which was a part of an embedded system based on DSP (TMS320DM642, TI Inc.) that we had previously developed. To avoid the effect of illegibility, the visual angle of main objects in these images was identical in principle. In Fig. 1, the last row consists of the simple scene images, the others are object images, and all of them have $288 \times 288$ pixels.

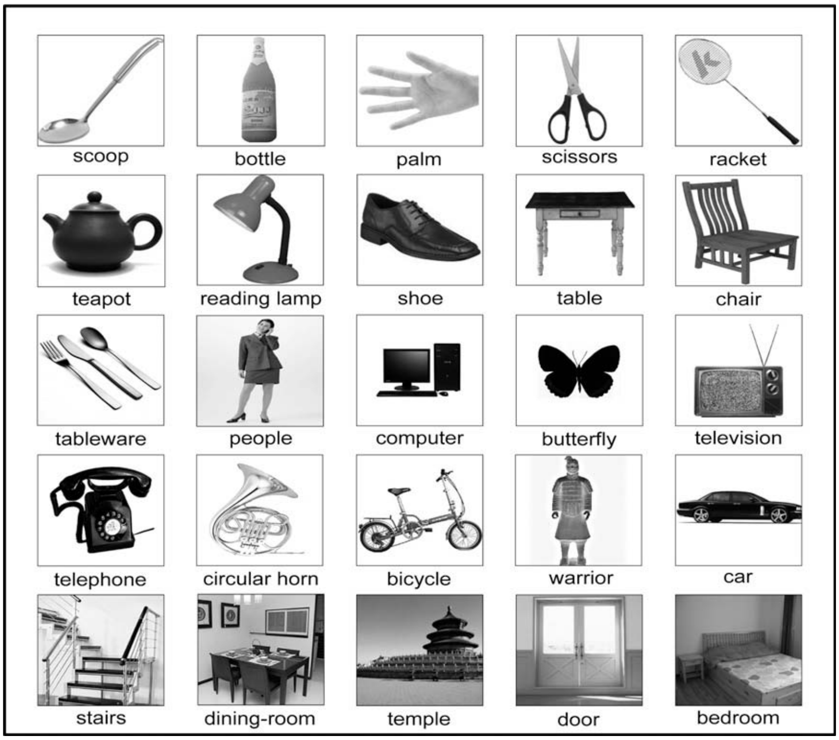


**Fig. 1.** The original object and scene images: the upper four rows contain objects and the last row contains the simple scenes.