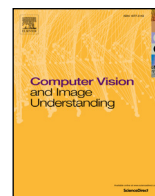




Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Semantic labeling for prosthetic vision

Lachlan Horne^{a,b,*}, Jose Alvarez^{a,b}, Chris McCarthy^{a,c}, Mathieu Salzmann^{a,b}, Nick Barnes^{a,b}^a NICTA, Tower A, 7 London Circuit, Canberra ACT 2600, Australia^b College of Engineering and Computer Science, Australian National University, ACT 0200, Australia^c Swinburne University of Technology, School of Software and Electrical Engineering, EN Building, John St, Hawthorn, Victoria, 3122, Australia

ARTICLE INFO

Article history:

Received 16 April 2015

Accepted 25 February 2016

Available online xxx

Keywords:

Prosthetic vision

Egocentric vision

Semantic labeling

ABSTRACT

Current and near-term implantable prosthetic vision systems offer the potential to restore some visual function, but suffer from limited resolution and dynamic range of induced visual percepts. This can make navigating complex environments difficult for users. We introduce semantic labeling as a technique to improve navigation outcomes for prosthetic vision users. We produce a novel egocentric vision dataset to demonstrate how semantic labeling can be applied to this problem. We also improve the speed of semantic labeling with sparse computation of unary potentials, enabling its use in real-time wearable assistive devices. We use simulated prosthetic vision to demonstrate the results of our technique. Our approach allows a prosthetic vision system to selectively highlight specific classes of objects in the user's field of view, improving the user's situational awareness.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Low or impaired vision is a common cause of disability, with prevalence rates estimated between 2.7% and 5.8% [1–3]. This represents a global health burden of nearly US\$3 trillion [4]. Those with low vision report reduced independence and social function, resulting in lower overall quality of life [5]. Prosthetic vision systems are a type of therapeutic device which aim to enable or improve key functional abilities in users with low vision, such as face recognition, reading and orientation and mobility [6,7]. In general, these devices attempt to replace the function of parts of the human visual system, which may be affected by disease or injury, by providing artificial stimuli to the user based on artificial sensory input [8].

There have been two large scale multi-centre chronic human trials of implantable visual prosthetic devices: The 60 electrode Argus II device [9], and the approximately 1500 electrode alpha-IMS device consisting of microphotodiodes implanted on the retina [10]. In addition, trials are underway by Pixium Vision SA (ClinicalTrials.gov identifier: NCT01864486) and a two year trial has recently been completed by Bionic Vision Australia [11].

The current state-of-the-art in prosthetic vision is limited in terms of functional outcomes for users. For example, the best visual acuity reported from the Argus II retinal prosthesis system is

20/1260 with a 20 degree field of view [9]. This is still well below the 20/200 threshold at which a person is considered legally blind in the United States¹. The number of discrete levels of stimulation intensity that a user can perceive is also limited. In retinal implants, up to ten levels may be discernable [12]. This limits users' ability to perceive contrast, and in systems where image processing is minimal, functional tests are usually performed in high contrast environments [13,14].

With these limitations, users often cannot interpret complex, uncontrolled scenes. However, computer vision and image processing techniques can be used in some systems to improve functional outcomes [11]. In this paper, we use *semantic labeling* techniques, which label each pixel in an image with a semantic category, to produce high-contrast stimuli from natural images. Stimuli based on semantic content can allow a user to distinguish objects in their field of view. This has the potential to improve functional outcomes for prosthetic vision users, by enabling navigation in complex environments. We use simulated prosthetic vision techniques, as in [15], to evaluate the resulting stimuli. An example is shown in Fig. 1.

We make three significant contributions in this paper:

- We apply semantic labeling to the problem of navigation using prosthetic vision. To our knowledge, this is the first application of semantic labeling in prosthetic vision. We show how semantic pixel labels can be converted to stimulation values so that important semantic distinctions can be made by the user. We

* Corresponding author at: NICTA, Tower A, 7 London Circuit, Canberra ACT 2600, Australia.

E-mail address: lachlan.horne@nicta.com.au (L. Horne).

<http://dx.doi.org/10.1016/j.cviu.2016.02.015>

1077-3142/© 2016 Elsevier Inc. All rights reserved.

¹ 42 U.S.C. §416(i)(1)(B) (Supp. IV 1986)

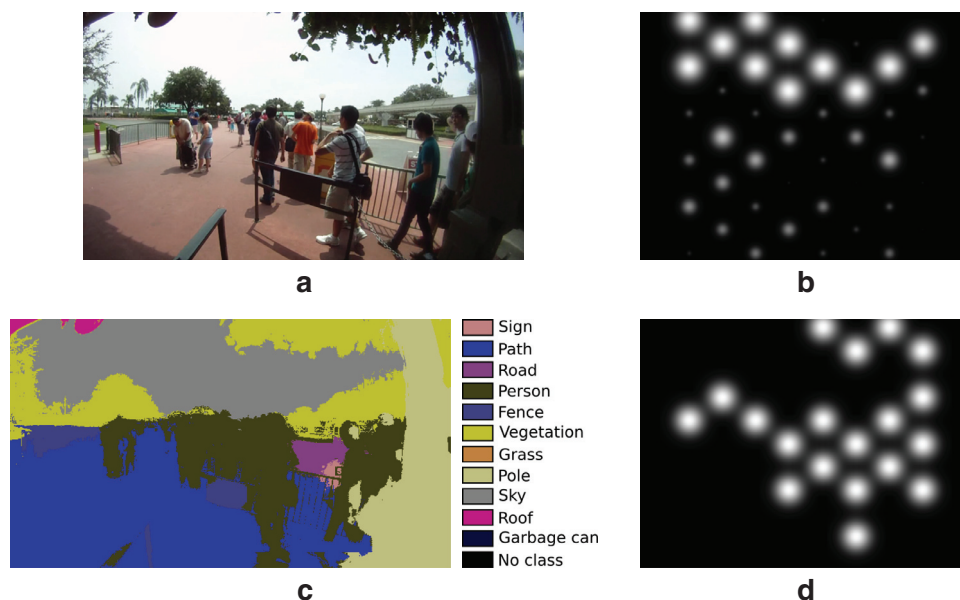


Fig. 1. Current prosthetic vision systems have limited ability to reproduce images with detail required for understanding an environment. Given an input image from a camera (a), simply downsampling image intensity values produces a visual percept that may be difficult to interpret, as our simulated prosthetic vision system described in Section 3 shows (b). Semantic labeling allows a computer system to produce a semantic representation of the scene (c), which can then be converted into a clearer percept (d). In this example we highlight potential obstacles, and dark regions show a safe path to walk.

demonstrate this in a range of navigation scenarios using simulated prosthetic vision.

- We introduce a new dataset for semantic labeling in egocentric vision. We provide semantic labels for 34 frames of the First-Person Social Interactions Dataset [16]. To our knowledge, no existing public egocentric vision datasets include pixel-wise semantic ground truth, and this represents a new and challenging application area for semantic labeling.
- We present a fast semantic labeling system which allows near real-time performance (faster than 1 Hz) for use in embedded devices. We improve upon our approach in [17] with a more accurate edge detector for predicting class boundaries, and a new cross-validation method to find selection parameters. We show improved and more detailed results on the CamVid [18,19], KITTI [20–22] and Stanford Background [23] semantic labeling benchmarks.

The paper is structured as follows. In Section 2, we give a review of previous work in vision processing for prosthetic vision, and semantic labeling. In Section 3, we present our framework for the application of semantic labeling in prosthetic vision. We highlight the need for improved vision processing techniques, and explain our model for incorporating semantic labeling. In Section 4, we present our fast semantic labeling system based on sparsely computed unary potentials. This enables the use of semantic labeling in a real-time embedded system, such as a prosthetic vision system.

In Section 5, we introduce our new dataset of manually labelled egocentric video. This allows us to qualitatively evaluate our proposed system in a realistic application scenario, using only images from a head-mounted camera in a dynamic environment. In Section 6, we measure the performance of our fast semantic labeling system on publicly available benchmark datasets. Then, we qualitatively demonstrate our proposed system, incorporating egocentric video, semantic labeling, and simulated prosthetic vision output. We conclude in Section 7.

2. Related work

2.1. Vision processing for prosthetic vision

Most current prosthetic vision systems use electrical stimulation to induce *phosphenes*, which are a sensation of light produced by stimulation other than light [24]. We focus on our work as part of an Australian government funded consortium, Bionic Vision Australia, which has recently conducted human trials showing that suprachoroidal implants are a viable technology for producing visual percepts [11]. Since this and similar systems use an external computer to process images from a camera to produce a stimulation pattern which is then fed to the implant, it is feasible to incorporate more complex computer vision techniques [25].

The potential for computer vision systems to improve functional outcomes in prosthetic vision users is often explored in the literature using simulated prosthetic vision. Simulation has also been used to predict functional outcomes in future devices [26,27]. Simulation allows a wide range of computer vision systems to be developed and tested without requiring implanted devices [6].

The use of abstract stimuli, that is, phosphene patterns not directly representing a downsampled version of a camera image, has been tested in the literature. Some approaches modify the field of view to aid the identification of objects. [28] uses zooming the field of view which is then communicated to the implant to aid in facial recognition, and [29] uses segmentation and zooming to aid in reading signs. For navigation, typically a wide field of view is used, and image processing compensates for poor acuity in this case. Visual saliency [30,31] can be used to highlight potential obstacles or objects of interest. Depth information, from stereo or RGBD cameras, can be used to detect obstacles [32]. [33] uses edges in depth images as well as face detection to produce a simplified high-contrast representation of the environment. Depth-based approaches have been tested using simulation [15,34].

Download English Version:

<https://daneshyari.com/en/article/6937574>

Download Persian Version:

<https://daneshyari.com/article/6937574>

[Daneshyari.com](https://daneshyari.com)