

A Salient Information Processing System for Bionic Eye with Application to Obstacle Avoidance

Ashley Stacey, Yi Li, and Nick Barnes

Abstract—In this paper we present a visual processing system for bionic eye with a focus on obstacle avoidance. Bionic eye aims at restoring the sense of vision to people living with blindness and low vision. However, current hardware implant technology limits the image resolution of the electrical stimulation device to be very low (e.g., 100 electrode arrays, which is approx. 12×9 pixels). Therefore, we need a visual processing unit that extracts salient information in an unknown environment for assisting patients in daily tasks such as obstacle avoidance. We implemented a fully portable system that includes a camera for capturing videos, a laptop for processing information using a state-of-the-art saliency detection algorithm, and a head-mounted display to visualize results. The experimental environment consists of a number of objects, such as shoes, boxes, and foot stands, on a textured ground plane. Our results show that the system efficiently processes the images, effectively identifies the obstacles, and eventually provides useful information for obstacle avoidance.

I. INTRODUCTION

Blindness is one of the most debilitating conditions that severely damages human's perception capability. For instance, there were 50,000 people legally blind in Australia in 2004, with numbers expected to increase to 87,000 by 2024 with the aging population [5].

There are many useful devices on the market to assist individuals with impaired vision. However, there is a lack of the implant systems that directly facilitate the human visual capability. In the past years, the fastest-growing biotechnology allows us to implant bionic devices to human patients, and further facilitate the patients to control their damaged sensory organs. Therefore, bionic eyes were recently proposed by a number of research institutes to restore human vision system.

A bionic eye consists of a camera that captures realtime videos and a microchip implanted in the retina. Certainly, the implant hardware is the most important component in the bionic eye, and the clinical surgery is the most critical process towards the success of these bionic eyes. However, current hardware technology limits the image resolution of the implant device to be extremely low. A 100 stimulation electrode array implant, which amounts to approximately 12×9 pixel resolution, would be the most reasonable and affordable device in the market for a number of years. Therefore, it necessitates a visual processing unit is to provide useful information for the microchip.

Ashley Stacey is with College of Engineering and Computer Science, Australian National University, Canberra, ACT Australia 2601. Email: ashley.stacey@anu.edu.au

Yi Li and Nick Barnes are with National Information and Communication Technology Australia (NICTA) and College of Engineering and Computer Science at the Australian National University, Canberra, ACT Australia 2601. Email: {yi.li, nick.barnes}@nicta.com.au



Fig. 1. Our visual processing system for extracting salient obstacles in images. On top of the helmet is a camera that captures videos that are processed by a laptop on the backpack. The results are visualized using the Emagin z800 head mounted display.

Among all the major considerations of bionic vision systems, navigation is a major milestone in the development process. Navigating in unknown environments by avoiding obstacles is very important in everyday life. However, this task becomes very challenging for bionic eye because a large amount of information is lost during the information processing and visualization in such a low resolution implant.

To assist bionic eye users to navigate in such a low resolution device, we must extract useful information from the environment to display. In our system, we attempt to use the salient information to identify objects in an unknown environment. Saliency has been studied for a decade in computer vision research. Traditionally, this is considered as one of the fundamentals in the biologically-inspired vision. Most existing methods model this as a bottom up process [9]. Low level visual cues, such as color and filter outputs, are combined together to create a higher level aggregation.

In our experiments, we placed a number of objects with different shapes, colors, and textures to simulate the obstacle avoidance process in an indoor environment. Our test environment ([1]) include white laced curtains and textured floor. The goal of the visual processing system is to identify the salient obstacles in the images captured by cameras.

Fig. 1 shows the implementation of our system. This portable system includes a camera (on top of the helmet), a laptop (in the backpack), and a head mounted display. The system is fully powered by the batteries in the backpack.

This paper is organized as follows: Sec. II discusses a few related topics, Sec. III presents our information processing procedure, Sec. IV provides experimental examples that demonstrate the robustness of the system in an indoor environment, and Sec. V concludes the paper.

TABLE I
SYSTEM CONFIGURATION.

Laptop	2.2 GHz Dual Core CPU with 2Gb memory
HMD	Emagin Z800 Visor
Webcam	Microsoft High Definition LifeCam.
System	Ubuntu 10.04
Dev Platform	Eclipse and OpenCV

II. RELATED WORK

A number of bionic eye systems are being developed world wide ([15], [7], [14]). These groups primarily focus on the implant modules. In our project, we attempt to develop novel visual processing systems to assist the implant module based on state of the art computer vision algorithms. For instance, extracting salient information is very helpful for low resolution device.

Computing salience is a major component of the areas of the biologically inspired computer vision. The surrounding world contains a tremendous amount of visual information. This is often regarded as a mediating mechanism [10] involving competition between different aspects of the scene.

Based on the work of Koch [3], Itti proposed a saliency-based visual attention model for scene analysis [9]. In this work, visual input is first decomposed into a set of topographic feature maps which all feed into a master “saliency map” in a bottom-up manner. This method has been proved to be an elegant map from biological theories to computational implementation. However, this approach is unable to match the eye tracking data well [11].

By considering the attention process as a signal processing problem, it is natural to use filters as simulation of the psychological mechanism [13]. For instance, subband filters [12] are widely used for this purpose. Bruce *et al.* [2] attempted to use information theory to define visual saliency and the fixation behavior. Other researchers ([8], [16]) argue that natural statistics should play an important role in the attention process.

Judd *et al.* [11] and Einhuser *et al.* [4] recently suggested that this attention problem should be tackled by taking both the low level features and the high level detectors into account. Human eye tracking data were then used for the training. In their work, human, face, and car detection were used to improve predicting human fixations. However, this method is computationally expensive in bionic eye.

III. SYSTEM IMPLEMENTATION

In this section, we first present the system configuration, then we present the flow chart of the salience algorithm, and finally we describe our processing procedures.

A. System requirement

As stated in the introduction, our system consists of three major components. This configuration is listed in Table I.

B. Saliency algorithm

Fig. 2 briefly describes the computation steps of the salience algorithm proposed by [6]. It first computes the

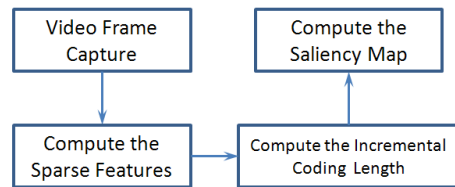


Fig. 2. Flowchart of the saliency extraction algorithm.

sparse features of the images, which are the linear weights of the sparse coding basis functions, and then uses the incremental coding length to compute the saliency.

We implemented this algorithm in C++ using OpenCV¹. The processing speed is approximate 0.2 second per frame. The profiler shows that the computation of the sparse features takes 80% of the total runtime.

C. Processing procedures

To process the images captured by the webcam, we perform the following operations.

- 1) Resize the image to 360×240 (Fig. 3, first row);
- 2) Compute the saliency map (Fig. 3, second row);
- 3) Binarize the saliency map to create the foreground image (Fig. 3, third row);
- 4) Resize the foreground image and the saliency map to 12×9 , and mask the saliency map using the foreground image in the reduced resolution.
- 5) Display the masked salience map (Fig. 3, forth row).

Our results are low resolution images that contain only foreground objects. Each pixel is proportion to the probability of the salience (importance) of being a foreground object.

IV. EXPERIMENT

In this section, we show that our system is capable of processing images and extracting salient foreground objects effectively. First, we demonstrate the effectiveness of the system by showing results for the objects in videos. Second, we compute the accuracy of the detection module and show the robustness of the system in an indoor environment.

We choose 12 objects that are commonly used in home/indoor environments every day. As listed in Fig 3, these objects are of different shape, texture, color, and other physical properties. All objects were placed in the experimental environment described in [1]. The subjects carried the system proposed in Sec. III in the environment, which captures videos and processes information.

The objects were placed on the ground. This is very challenging because the ground has textures with similar colors, and the areas between wall (curtain in our case) and the ground is challenging to handle. Please note that in this case there is no significant difference between moving objects and moving the camera, because the algorithm does not take optical flow into account.

In total, we captured 10 videos with 600 images per video. The capturing frame rate of the webcam is 24 frame

¹<http://opencv.willowgarage.com/wiki/>

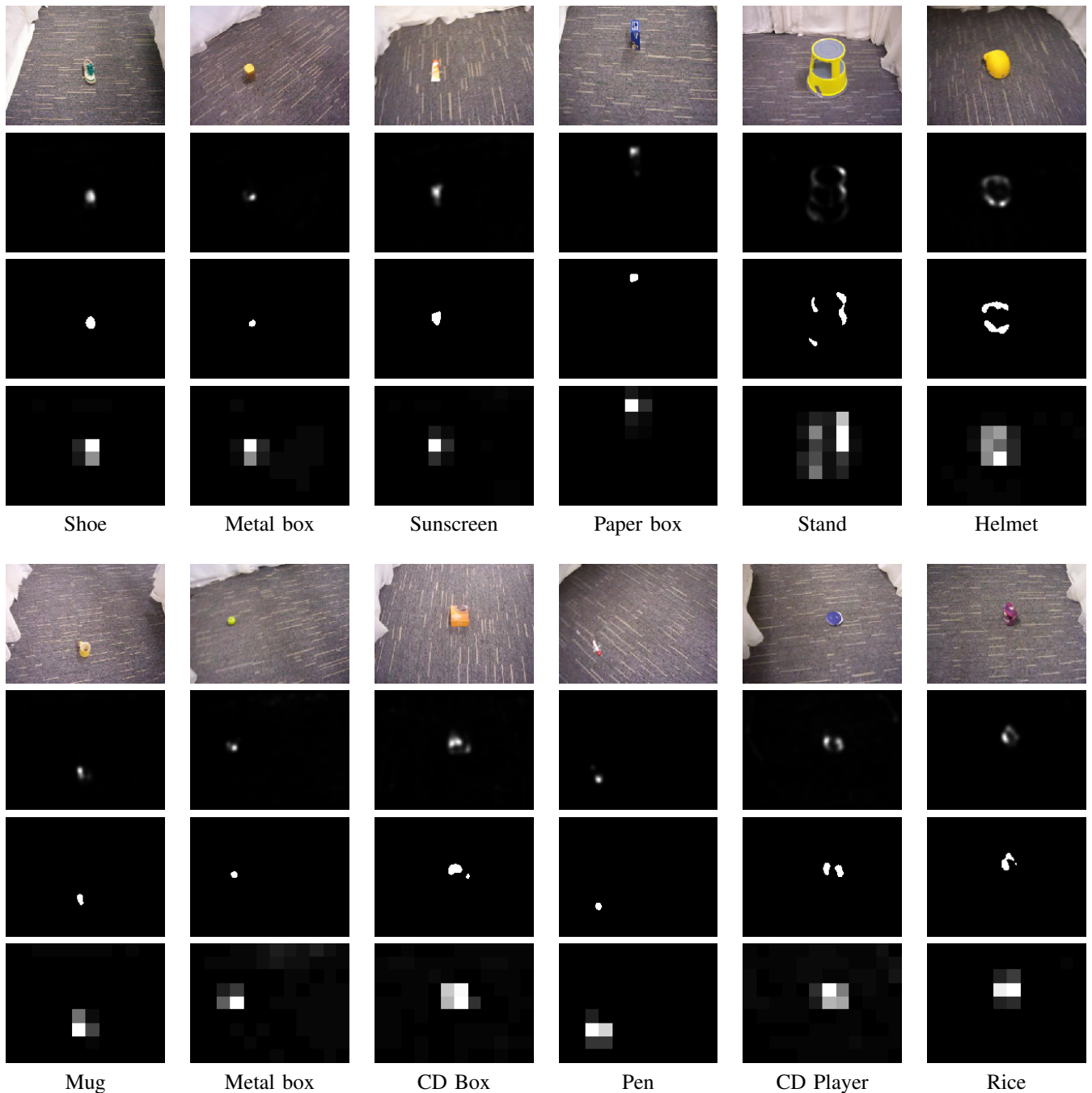


Fig. 3. Experimental setups and examples. We choose 12 common objects, which include a shoe, a metal box, a sunscreen, a paper box, a stand, a helmet, a mug, a metal box, a CD box, a pen, a CD player, and a bag of rice. The objects were placed on the textured ground. The first/fifth row: captured images; The second/sixth row: saliency maps; The third/seventh row: foreground images; The four/eighth row: low resolution display;

per second. However, we only process one every 4 frames because the time for processing each image is 0.2 second.

A. Example results

We show the experimental configuration and the example images of the objects in Fig. 3. Since the images were captured while the subjects were walking through the environment, the images were taken from different angles and directions, and motion blur is unavoidable. The texture on the carpet posed additional difficulties because the sizes, orientations, and colors of the textured strips were different.

Fig. 3 shows that our system successfully captured the salient objects in the scene. The low resolution display (forth/eighth row in Fig. 3) successfully handles the textures of background and the intersection regions between walls and the ground, and eventually effectively pinpoints the locations of the obstacles on the ground plane.

B. Results on videos

Ideally, the low resolution saliency must be consistent in the same video. We show such an example in this section.

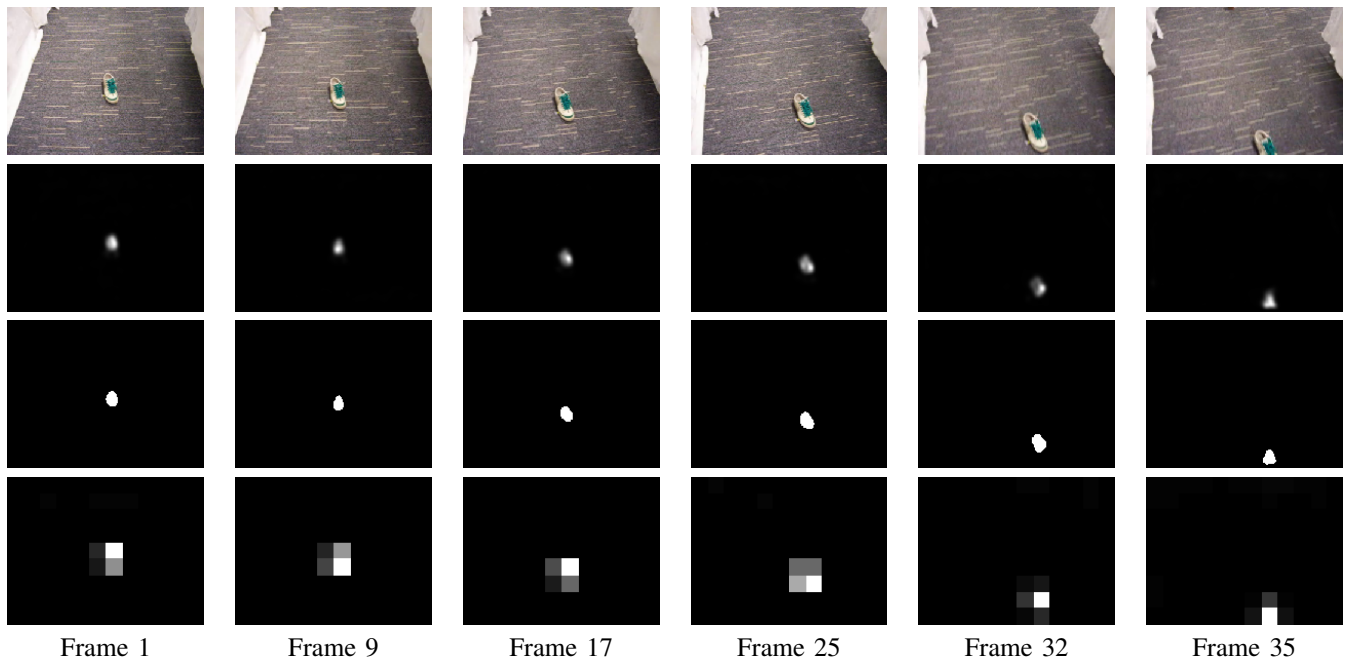


Fig. 4. This example shows a number of frames in the same video.

Fig. 4 displays a number of frames in the same video when the subject is passing by a shoe. One can see that the shoe can reliably be detected for all the frames and the system performance is robust to the viewpoint change.

C. Robustness analysis

We compute the robustness of the system in this section. Define the hit rate of the algorithm as the number of the successful detected objects over the total number of frame, we achieved approx. 96% accuracy.

The failures are mostly from the images that do not contain images (Fig 5). In this case, the most salient background regions may be mistakenly identified as foreground images and created false positive detections.

V. CONCLUSION AND DISCUSSION

We present a visual processing system for bionic eye in this paper. This fully portable system includes a camera, a laptop, a state-of-the-art saliency detection algorithm, and a head-mounted display. Our results show that the system effectively identifies these obstacles and eventually provides useful information for navigation tasks. Our algorithm is designed for the 100 electrode array implant, and can be implemented efficiently using hardware.

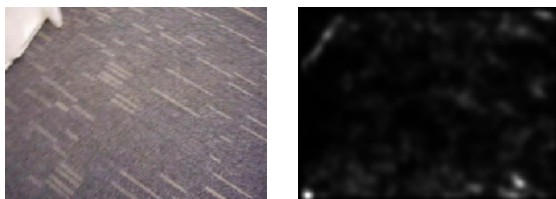


Fig. 5. A failure example.

In the future, a 1,000 stimulation array will be in the market. This enables the users to identify regions in details. Our algorithm would serve as an effective component for finding salient regions for image zooming and stabilization.

REFERENCES

- [1] Nick Barnes, Paulette Lieby, Hugh Dennet, Janine Walker, Chris McCarthy, Nianjun Liu, and Yi Li. Investigating the role of single-viewpoint depth data in visually-guided mobility. In *Vision Science Society Annual Meeting 2011*, 2011.
- [2] N. D. B. Bruce and J. K. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):1–24, 3.
- [3] C.Koch and S.Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.
- [4] Wolfgang Einhäuser, Merrielle Spain, and Pietro Perona. Objects predict fixations better than early saliency. *Journal of vision*, 8(14), 2008.
- [5] Centre for Eye Research Australia. Eye research australia clear insight: The economic impact and cost of vision loss in australia.
- [6] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *CVPR 2007*, 0:1–8, 2007.
- [7] <http://retina-implant.de/en/default.aspx>.
- [8] L. Itti and P. Baldi. Bayesian surprise attracts human attention. *NIPS*, 19:547–554, 2006.
- [9] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews, Neuroscience*, 2:194–203, 2001.
- [10] W. James. *The Principles of Psychology*. Holt, New York, 1890.
- [11] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. *ICCV*, 2009.
- [12] Javier Portilla and Eero P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40:49–71, 2000.
- [13] J. M. Wolfe and K. Cave. *Deploying visual attention: The guided search model*. John Wiley And Sons Ltd., 1990.
- [14] K.J. Wu, C. Zhang, W.C. Huang, L.M. Li, and Q.S. Ren. Current research of c-sight visual prosthesis for the blind. In *EMBC 2010*, 2011.
- [15] www.2-sight.com.
- [16] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7):1–20, 2008.