

# Detecting Objects and Obstacles for Visually Impaired Individuals using Visual Saliency

Benoît Deville  
Benoit.Deville@unige.ch  
Computer Science Department,  
University of Geneva  
Route de Drize 7  
1211 Carouge  
Switzerland  
+41223790177

Guido Bologna  
Guido.Bologna@unige.ch  
Computer Science Department,  
University of Geneva  
Route de Drize 7  
1211 Carouge  
Switzerland  
+41223790151

Thierry Pun  
Thierry.Pun@unige.ch  
Computer Science Department,  
University of Geneva  
Route de Drize 7  
1211 Carouge  
Switzerland  
+41223790153

## ABSTRACT

In this demo, we present the detection module of the See CoLoR (Seeing Colors with an Orchestra) mobility aid for visually impaired persons. This module points out areas that present either particular interest or potential threat. In order to detect object and obstacles, we propose a bottom-up approach based on visual saliency: objects that would attract the visual attention of a non-disabled individual are pointed out by the system as areas of interest for the user. The device uses a stereoscopic camera, a laptop, and standard headphones. Given the type of scene and/or scenario, specific feature maps are computed in order to indicate areas of interest in real-time. This demonstration shows that the module indicates objects and obstacles as accurately as a system using all available feature maps.

## Categories and Subject Descriptors

I.4.8 [Scene Analysis]: Color, Depth cues, Stereo; K.4.2 [Social Issues]: Assistive technologies for persons with disabilities, Handicapped persons/special needs.

## General Terms

Design, Experimentation, Human Factors.

**Keywords:** Visual impairment, Mobility aid, Visual saliency, Stereo, Video processing, Color.

## 1. INTRODUCTION

In order to orientate themselves, most visually impaired individuals (i.e. blind or with low vision) use a white cane, a guide dog, a clicking sound or a combination of those methods. However, some simple tasks are still issues for visually impaired people, like finding small objects, avoiding obstacles on their walkway, or finding the entry of a building.

So far, a large number of tools have been introduced to allow visually impaired individuals to explore and navigate their environment. The most usual ones are the white cane and the guide dog, but more technological systems have been investigated over the last decades.

---

Demo videos: <http://cvml.unige.ch/doku.php/members/bdeville>

Copyright is held by the author/owner(s).  
*ASSETS'10*, October 25–27, 2010, Orlando, Florida, USA.  
ACM 978-1-60558-881-0/10/10.

See CoLoR [2] is a mobility aid<sup>1</sup> designed to help visually impaired people to obtain information about an unknown environment by transforming colored pixels into musical instrument sounds. A number of experiments have demonstrated that simple tasks like following a colored line, finding a colored object, or matching objects according to their color, can be easily performed using this synesthetic approach. However, users have expressed additional needs, like finding small objects on a table, preventing stumbling into aerial obstacles, or finding a doorway into a building. The detection module we present is based on visual saliency. Specifically, objects that would attract the visual attention would be pointed out as areas of interest for the user. We present here a preliminary answer to this need.

## 2. ASSISTIVE DEVICES

The most intuitive way to help persons with visual impairment finding their way in an unknown environment is the use of GPS-based systems. Nowadays, some mobile phones offer this characteristic, and some devices have been adapted to this need, like the Personal Guidance System [5] or the BrailleNote GPS [7]. However, position accuracy is still an issue.

Another promising method consists of creating a 3D model of the environment around the user, or at least a part of it. It can be used, for instance, for detecting steps [8].

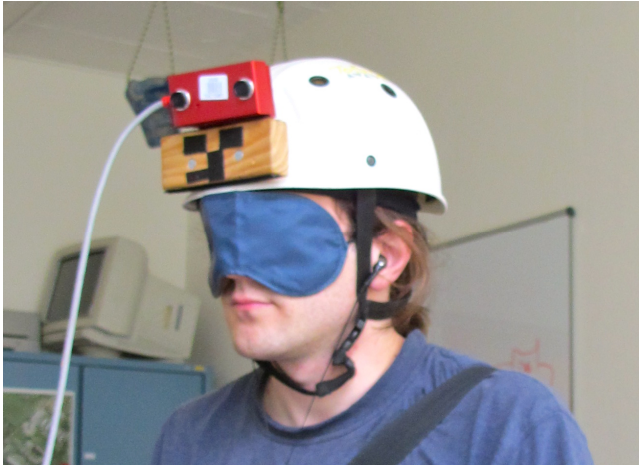
Finally, the most investigated area is probably the sensorial substitution approach, where touch and/or audition are employed. Tactile displays are amongst the oldest devices [1]. They usually reproduce a pattern on the user's skin with small pins or using piezoelectricity. An auditory display can be a sonar or any device that transform an image of the environment into sounds, like the vOIce [6]. Most auditory displays require the sounds to be permanently played and some can be felt as annoying by the user. Actually, as shown by a study performed in 2003 [3], visually impaired users give more credit to speech-based devices. This is why the detection module describes information about objects and obstacles in oral language.

## 3. DESIGN

### 3.1 Hardware

Our prototype consists of a stereoscopic camera connected to a common laptop through an IEEE 1394 Firewire interface. The camera used with this prototype is a Videre STOC-6CM-C-

MINI<sup>2</sup>. It weighs less than 200g and computes the disparity map, i.e. the distance information, in real-time (~25Hz) on a dedicated chip, on 640x480 color images. It is positioned on a protective helmet so that the user does not need to bother about its position when it operates. A dedicated battery powers the camera since laptop Firewire interfaces do not provide electrical energy. The user wears a small laptop case where are placed both the laptop and the camera battery, as illustrated on Figure 1. Any standard headphone is suitable for this prototype.



**Figure 1.** A blindfolded user wearing the SeeColOr prototype.

### 3.2 Methods and software

The whole detection module has been programmed in C++. In fact, the prototype needs to react as close as possible to real-time, and the STOC camera can be driven through a C++ library. Furthermore, the saliency detection part is based on the iLab Neuromorphic Vision C++ Toolkit (iNVT)<sup>3</sup>, a visual saliency library based on Itti *et al.*'s work [4]. We still had to adapt the iNVT so that both color and depth information from the stereo camera could be used as inputs for the visual saliency process. The alert messages are composed of vocal sentences, recorded in quiet conditions.

The module works as follows.

1. The user defines its objective, e.g. finding an object on a table, finding a door, or detecting an obstacle.
2. The system automatically selects the set of feature maps that have to be computed.
3. For each frame, a center-surround difference is applied on each feature map to provide conspicuity maps.
4. Conspicuity maps are combined into a final saliency map. Peaks of this map are focuses of attention (FOA).
5. When a FOA reaches a given threshold (i.e. the highest peak over a given number of frames), the user

is informed of the alert with the object/obstacle's global position through an explicit vocal message.

## 4. PRELIMINARY RESULTS

Initial tests using the detection module system have been performed on a set of videos shot in stereo. Each video clip depicts a situation where visually impaired individuals might have trouble finding their way or the desired object. In fact small and/or flat objects, aerial or thin obstacles are easily missed with a white cane or by a guide dog. These tests show some very interesting results. First, the computing time for FOA on each frame approaches real-time (about 200ms), which allows fairly fast reactivity to sudden changes. Moreover, the approach is robust since removing some feature maps did not affect the accuracy or the relevance of pointed areas (the desired objects/obstacles).

Real-life experiments with both blindfolded and visually impaired participants will now be performed in order to obtain qualitative results. Users will for example be asked to follow a red line as done in a previous study, and to avoid natural obstacles that will be detected by the device. Another experiment will need them to enter a room, and find a small object on a table. Both experiments will be based on scenarios depicted in the set of video clips used for the initial tests.

## 5. REFERENCES

- [1] Bach-y-Rita, P., Collins, C., Saunders, F., White, B. and Scadden, L. 1969. Vision substitution by tactile image projection. *Nature* 221 (1969), 963-964.
- [2] Bologna, G., Deville, B. and Pun T. 2009. On the use of the auditory pathway to represent image scenes in real-time. *Neurocomputing* 72, 839-849.
- [3] Golledge, R., Marston, J., Loomis, J. and Klatzky, R. 2004. Stated Preferences for Components of Personal Guidance System for Nonvisual Navigation. *Journal of Visual Impairment & Blindness* 98 (March 2004), 135-147.
- [4] Itti, L., Koch, C. 2001. Computational Modelling of Visual Attention. *Nature Reviews Neuroscience* 2 (March 2001), 194-203.
- [5] Loomis, J., Golledge, R., & Klatzky, R. 2001. GPS-based navigation systems for the visually impaired. In: Barfield W, Caudell T (eds), *Fundamentals of Wearable Computers and Augmented Reality*, 429-446.
- [6] Meijer, P. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39 (February 1992), 112-121.
- [7] Ponchilla, P., Rak, E., Freelan, A. and LaGrow, S. 2007. Accessible GPS: Reorientation and Target Location Among Users with Visual Impairments. *Journal of Visual Impairment & Blindness* 101 (July 2007), 389-401.
- [8] Pradeep, V., Medioni, G. and Weiland, J. 2008. Piecewise Planar Modeling for Step Detection using Stereo Vision. *Workshop on Computer Vision for Applications for the Visually Impaired, ECCV'2008* (Marseille, France, October 18, 2008).

<sup>2</sup> Videre Design: <http://www.videredesign.com>

<sup>3</sup> iNVT: <http://ilab.usc.edu/toolkit/>