Supporting Blind Navigation using Depth Sensing and Sonification

Michael Brock
School of Computer Science
University of St Andrews
St Andrews, United Kingdom
mob5@st-andrews.ac.uk

Per Ola Kristensson
School of Computer Science
University of St Andrews
St Andrews, United Kingdom
pok@st-andrews.ac.uk

Figure 1: A blind-folded user equipped with the navigation support system.

Abstract

We present a system designed to help blind people navigate around obstacles. Our system perceives the environment in front of the user using a depth camera (a Microsoft Kinect). The system identifies nearby structures from the depth map and uses sonification to convey obstacle information to the user. The system has undergone a formative evaluation involving eight blind-folded participants and one blind participant. We found that our system can be learned within minutes and that participants can successfully navigate through an obstacle course with few collisions.

Author Keywords

Blind navigation, accessibility

ACM Classification Keywords

K.4.2 [Social Issues]: Assistive technologies for persons with disabilities.

Introduction

We present a new navigation aid for the blind that sonifies the surroundings of the user based on data gathered from a depth sensor (a Microsoft Kinect). Blind and visually impaired people heavily rely on guide dogs, their cane, and/or bystanders when navigating around unfamiliar places. Blind people generally have a good sense of

---

UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.
ACM 978-1-4503-2215-7/13/09.

http://dx.doi.org/10.1145/2494091.2494173

Figure 2: The raw RGB VGA image from the camera.

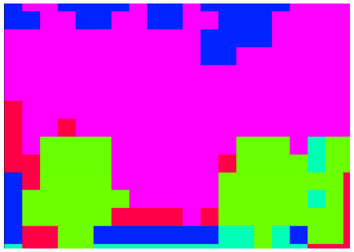

Figure 3: The depth map after down-sampling.

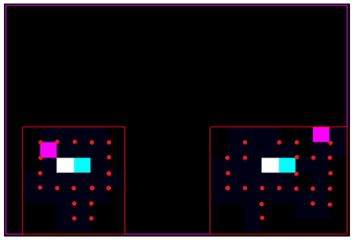

Figure 4: Detected structures to be sonified.

orientation and memory, allowing them to navigate without problems in familiar environments.

The cane is the standard navigation tool for the blind and it is effective due to its simplicity. However, the cane is also physically restricted to a limited range. The aim of our system is not to replace the cane but to *complement* it with depth sensor-based sonification to increase the effective "sensor range" for blind people navigating via a cane.

Depth cameras have been extensively used in a variety of applications, including obstacle detection, gesture recognition and environment mapping. However, using depth cameras in conjunction with sonification to design new navigational systems for blind users is still an underexplored area of research. Edwards and colleagues [2] provide haptic feedback via vibrotactile motors to convey the position and distance to an obstacle via the position and strength of vibrations. Another system is "the vOICe" [1], which uses a webcam to translate video data into a sound stream. The vOICe sonifies pixels by altering the volume depending on the brightness of the pixels, and by altering the pitch depending on the vertical position of bright pixels. The horizontal position is conveyed by panning the sound from left to right in a one second repeated cycle [1]. A recent study has revealed that the complexity of the vOICe required users to use it continuously for three weeks in order for them to effectively being able to locate 2D objects [4].

The design of our own system was inspired by a small minority of people having developed the skill of human echolocation. This allows them to navigate without additional navigation aids. This technique uses background noise or self-made clicks to sense obstacles.

System

Figure 1 shows a blind-folded user equipped with our portable system.

Basis for System Design

The design of the system was informed by two structured interviews. We interviewed two potential end-users before we began designing the system. The first interviewee was a second-year university student with 5% vision. The second interviewee was a first-year university student who became blind in his early childhood.

Major findings from the interviews include: (1) The blind interviewee could not navigate efficiently without an assistant in unfamiliar areas. (2) The interviewee with 5% vision stressed the desire to perceive colour while the interviewee who was blind since early childhood did not feel colour perception was important. (3) The interviewee with 5% vision stressed the importance of being able to customise the system. (4) The interviewee who was blind since early childhood stressed the importance of using subtle audio feedback because this person relied heavily on audio to perceive his surroundings (human echolocation).

Based on this research we concluded that our system did not necessarily require the ability to sonify colour. However, our system must be customisable to accommodate for different sound preferences, and the act of sonification must not over-power the user.

The blind interviewee navigated on his own indoors using human echolocation. This technique allowed him to stop in front of closed doors, walk through corridor passages while avoiding the walls, and identifying the corners where he needed turn.

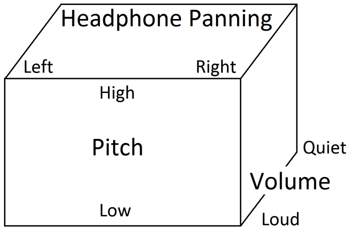

Figure 5: The conversion from 3D location to sound within a given sonification volume. The horizontal (x) position of an obstacle is encoded by the panning position. Obstacles on the left will thus be heard predominantly in the left headphone, and vice versa. The vertical (y) position is encoded by varying the pitch. A high pitch implies high elevation. Finally, the distance to the user (z) is encoded via volume. Obstacles closer to the user are louder than obstacles farther away.

For example, a small obstacle at the feet of the user would generate a loud, low pitch sound in both headphones. A distant hanging object to the right of the user would generate a faint, high pitch sound in the right headphone.

While demonstrating his navigation abilities he walked into a sofa, which was too low down for him to sense using echolocation. He concluded that an effective navigation aid for the blind must help avoid situations like the aforementioned, in which an object simply cannot be detected via a cane, blind dog or echolocation.

System Architecture
The system works by continuously carrying out the following steps:

1. Reading depth data from the Kinect sensor.
2. Down-sampling depth data into a low-resolution image.
3. Performing obstacle detection.
4. Sonifying obstacle information to the user's headphones.

The down-sampling reduces the resolution of the depth frames from 640×480 to 20×15 pixels. In the down-sampling process, pixels that are closer to the user are prioritised. Thereafter obstacles are detected by splitting the depth data into isolated structures at different depth levels using the marching squares algorithm [3].

Having identified obstacles at various depth levels, we use a SuperCollider sound synthesis server[1] for sonification. Our system uses a WAV file for sonification and modifies its pitch, panning position and volume. Each obstacle at the closest depth level is converted into a series of sounds, where the x and y components of the obstacle are averaged. The exact conversion from 3D coordinate to sound is illustrated in Figure 5. This system is novel because of its use of a depth sensor in conjunction with a simple world-to-sound mapping that allows blind people to navigate more confidently around unfamiliar areas.

Formative Evaluation

We have carried out two experiments in order to determine to what degree people could learn this system and to what extent the system allowed them to navigate around an obstacle course.

Blind-Folded Participants
For the first experiment 8 participants were recruited, none of which were blind or had a visual impairment. Their ages ranged from 18 to 25 (mean = 22, sd = 2), 3 were male and 5 were female. Participants were first informed about how the system worked. Once participants understood the basics of the system, a tutorial taught them how the sound mapping worked. The tutorial sonified and visualised objects of various sizes and positions in 3D space alongside a synthesised voice that described the position. This allowed participants to gain an understanding of how each of the sound variables were affected by an object's position and size.

With this understanding participants were presented with a 3×3 grid illustrating the possible location of objects they were about to hear. Participants were exposed 18 single sounds. After each sound, participants were instructed to locate the origin of the sound on the grid. They were also asked to state whether they believed the object was near or far. Participants controlled the playback of the sounds themselves and could therefore progress at their own speed. We used one sound from each of the grid's nine locations at a near and far distance, 18 sounds in total. Sounds were played in random order.

After this experiment, participants were blindfolded and given a Kinect and a laptop bag to strap around their

---

[1] `http://supercollider.sourceforge.net/`

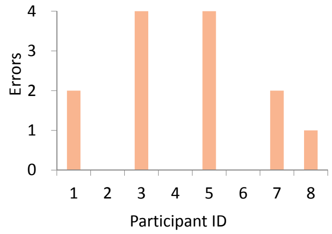

Figure 6: Number of obstacles each individual participant hit in the evaluation.

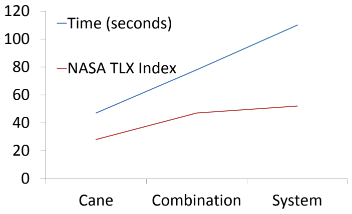

Figure 7: The results from the study with a blind participant. The NASA TLX Task Load Index is used to measure the blind participant's perceived workload during the task (lower is better).

neck. They were then guided to an obstacle course consisting of chairs, tables and walls forming a single path. Participants were instructed to walk along this path. When they arrived at the end, the experimenter turned them around and requested participants to walk back to the starting location.

Blind Participant
We recruited a single male blind university student who was 18 years old and had become blind at the age of two. He was given the same oral debrief and tutorial as the sighted participants but was allowed to test the system and modify the user settings. This was done because we wanted the system to be used as intended by blind people. Three different obstacle courses were setup, all using the same number of chairs and all using similar layouts. During the first trial the participant used his cane for navigation, during the second trial he used our system, and during the last trial he was allowed to use his cane and our system in combination.

Results
Our initial results are promising. The blind-folded participants all managed to navigate the obstacle course with few errors (see Figure 6). Our blind participant was faster when navigating with the cane (Figure 7). This was expected as this participant had years of training in using the cane. Encouragingly he also quickly learned how to use our system. Our results show that our method of sonifying 3D obstacles is effective. In addition, users are able to use the system after just a few minutes of practice.

Conclusions and Future Work

We have presented a system designed to help blind people navigate around obstacles using a depth camera. Our system sonifies nearby obstacles to the user. The system has undergone a formative evaluation involving eight blind-folded participants and one blind participant. The results show that our system can be learned within a few minutes and that participants can successfully navigate through an obstacle course with few collisions.

We conjecture our system could be improved by fusing data from multiple sensors. This could both improve the accuracy of the obstacle detection algorithm and increase the sensor range of our system. Further, in order to better understand the potential of our system and the nuances of this type of interaction, the system will need to be evaluated in realistic use-contexts in longitudinal experiments with blind people.

Acknowledgement

This work was supported by the Scottish Informatics and Computer Science Alliance (SICSA).

References

[1] Artificial vision for the totally blind. `http://www.artificialvision.com/`.
[2] Edwards, N., Rosenthal, J., Moberly, D., Lindsay, J., Blair, K., Krishna, S.; McDaniel, T., and Panchanathan, S. A pragmatic approach to the design and implementation of a vibrotactile belt and its applications. In *Proc. HAVE 2009*, IEEE Press (2009), 13–18.
[3] Lorensen, W. E., and Cline, H. E. Marching Cubes: A high resolution 3D surface construction algorithm. In *Proc. SIGGRAPH 1987*, ACM Press (1987), 163–169.
[4] Proulx, M., Stoerig, P., Ludowig, E., and Knoll, I. Seeing where through the ears: Effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution. *PLoS ONE 3*, 3 (2008), e1840.