

Electronic Travel Aid Based on Consumer Depth Devices to Avoid Moving Objects

Jee-Eun Lee

NCsoft Corporation, Seoul, South Korea

Yujin Kim

LOTTE.com Inc., Seoul, South Korea

Taejung Park

Visual Media Lab., Department of Digital Media
Duksung Women's University, Seoul, South Korea

Copyright © 2016 Jee-Eun Lee, Yujin Kim and Taejung Park. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

We present a prototype for an electronic travel aid (ETA) for visually impaired persons based on consumer depth sensors. To provide more helpful information, our ETA system detects the velocity of moving objects and helps the user to avoid them with proper audible warnings. To achieve the stated goal, our system captures and processes depth images in front of a user to remove noises and acquire a bounding box for each object. Then, it traces the objects' bounding boxes and their speeds and directions of movement in the captured depth images. According to the recognized directions and speeds, our system warns the user not only of immediately dangerous situations, but also of potential risks, via audio warnings and cautions.

Keywords: electronic travel aids, depth sensors, visually impaired, Kinect, Xtion

1 Introduction

Many people with weak vision or visual impairment find it very difficult to walk on a road with many moving objects. They usually rely on a white cane on

roads. Unfortunately, since the white cane can only detect low-profiled obstacles and patterns on the road, it often fails to guide a person safely and successfully due to vertical obstructions including traffic signs, safety mirrors, and billboards. (Fig. 1). Moreover, it makes it more difficult for the visually impaired to walk through crowded roads with many moving and static objects having various velocities.



Figure 1. Vertical obstacles that are hazards for the visually impaired person walking with only a white cane

To address these issues and help the visually impaired people move more conveniently, various electronic travel aids (ETAs) and robotic travel aids (RTAs) have been proposed. Most of these aim to overcome the limitations of the white cane and provide the user with more ability by applying global positioning system (GPS) [1], ultrasonic sensor [2], or stereo vision technology [3]. Although those ETA techniques with GPS features could help the user travel long distances, they often fail to work in buildings and are not suitable for shorter trips. Also, the ETAs based on ultrasonic or light sensors can detect the existence of obstacles, but would fail to detect smaller obstacles and would not provide the user with detailed information (e.g., the directions or speeds of the moving obstacles). Those with stereo vision require multiple (at least two) cameras and higher computing powers to analyze multiple two-dimensional videos and construct three-dimensional scene information, which would inevitably make the whole system heavier and reduce its battery life.

To overcome these limitations, we propose an ETA prototype that uses consumer depth images like those from Microsoft Kinect, ASUS Xtion, and Occipital Structure Sensor. Our ETA system detects the velocity of moving objects and helps the user to avoid them with proper audible warnings.

Recently, some ETA systems similar to ours have also adopted depth sensors. Oxford Smart Sunglasses [4] uses similar depth sensors to help people with weak vision recognize objects more easily. However, the purpose and aim of this device are different from those of ours; it depends on depth sensor images only to remove the details of a real view. Besides, those who have lost their sight entirely cannot use this device. Kinechesia [5] also uses the Kinect depth sensor and shares a very similar structure with our approach. However, this device cannot provide detailed information as ours does and only directs the user to three options: left, middle, and right. We have introduced overview of our device in a

conference [6], but not full details. We discuss the details fully in this paper. Table 2 summarizes some of the related ETA technologies and their limitations.

Table 1. Related technologies

| Name | Description | Limitations |
|--------------------------------|---|--|
| Oxford Smart Sunglasses [4] | These smart sunglasses rely on depth images, as does our method. By projecting abstract depth images, they help people with weak vision to recognize objects. | The purpose of this device is to remove details of a scene to help the people with weak vision to recognize the objects more easily. They are not useful for those who have completely lost their sight. |
| Kinesthesia [5] | Using a Kinect depth sensor, this device directs the user by vibration. | This device only indicates three options: left, middle, and right. It does not provide detailed information (e.g., the distance or velocity of the objects). |
| ETA based on stereo vision [7] | With two video cameras, this device analyzes and recognizes numbers, letters, and symbols on the streets (e.g., bus stops, billboards, etc.), and reads them aloud to the user. | This device focuses on providing visual information around the user in an audible way, without including safety information about moving objects. |

2 Proposed System

2.1 System Overview

Figure 2 shows the overview of our system. Our device captures depth images from depth sensors (e.g., Microsoft Kinect, ASUS Xtion, or Occipital Structure Sensor). These three sensors use similar technologies, but the Structure Sensor is the lightest and is smaller than the others are. In this regard, the Structure Sensor is more suitable for our purpose. Based on preliminary tests, we concluded that the distance from the user to the moving object should be 3 to 4 m to warn the user safely. Table 2 summarizes the specifications of the three depth sensors. Since all three of the sensors can be accessed by OpenNI SDK [7], there is little difference between them in the development of our system.

Table 2. Depth device information

| | Weight (g) | Size (L×W×H) mm | Range (cm) |
|------------------|------------|---------------------|------------|
| Kinect v1 | 1,360 | 304.8 × 76.2 × 63.5 | 40 - 450 |
| Xtion | 226.7 | 177.8 × 50.8 × 38.1 | 80 - 350 |
| Structure | 95 | 119.2 × 27.9 × 29 | 40 - 350+ |

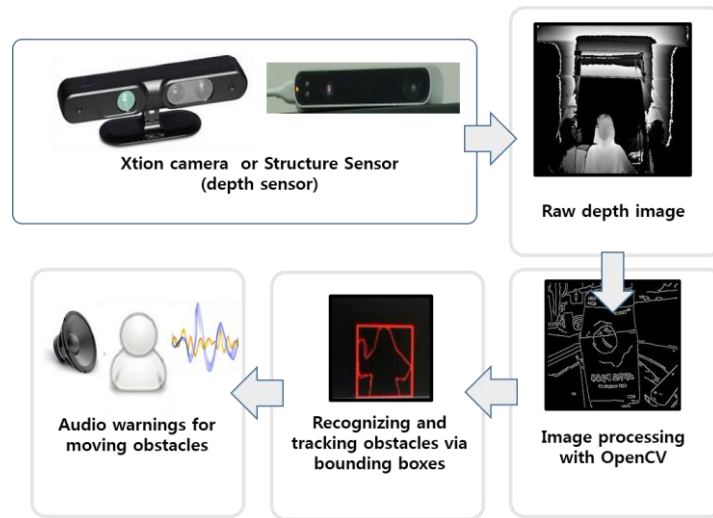


Figure 2. System overview of the proposed ETA device [6]

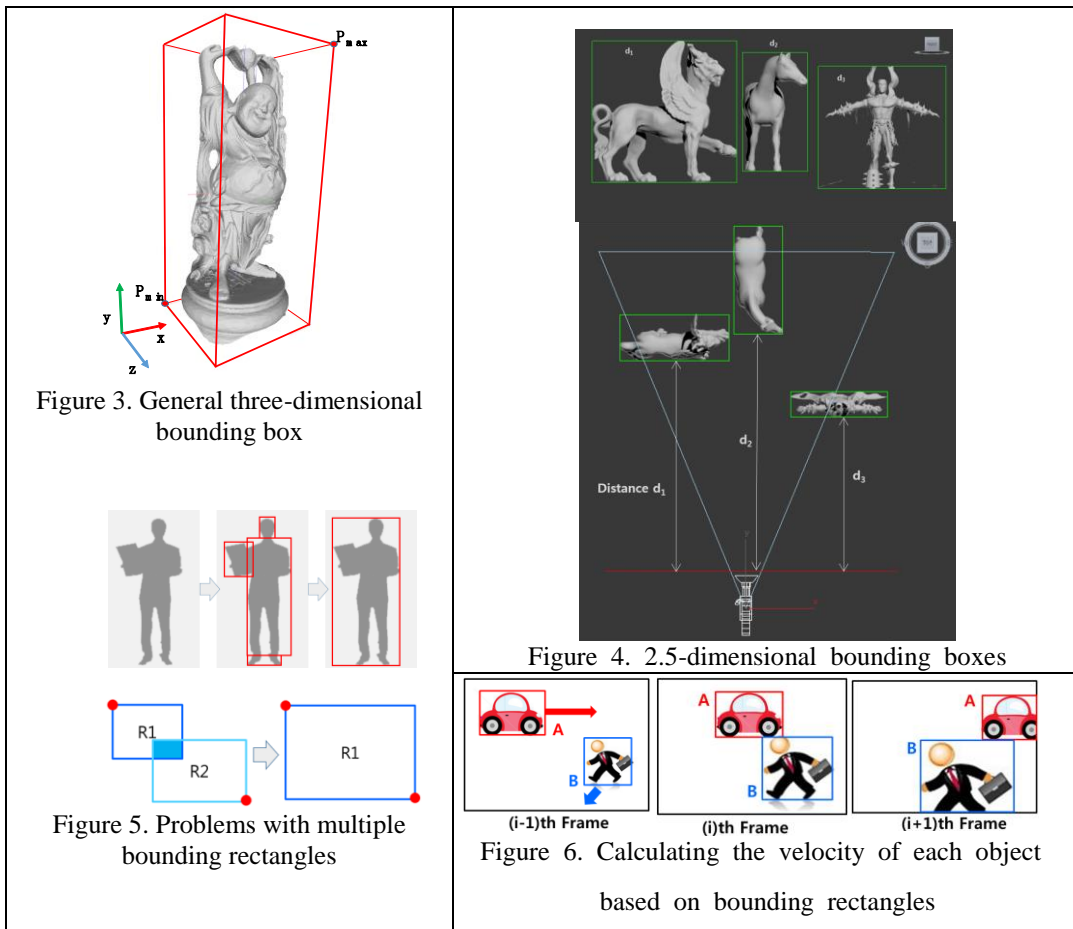
2.2 Processing Depth Images

In general, the depth images captured from the depth sensors have much noise. Therefore, we need to denoise the depth images to achieve robust results. For this goal, we apply OpenCV to the depth images in an embedded Linux environment based on Raspberry Pi [8].

We stabilize the depth images with a Gaussian blur filter followed by a Canny filter [9, 10] to construct bounding boxes that can recognize each object separately.

2.3 Detecting and Tracing Objects

In the next step, we construct bounding boxes with the stabilized images in the previous step. Our system recognizes and traces each object with the movement of its bounding box. In general three-dimensional cases, bounding boxes can be constructed by acquiring the maximum point $P_{\max} = (x_{\max}, y_{\max}, z_{\max})$ and minimum point $P_{\min} = (x_{\min}, y_{\min}, z_{\min})$ (see Fig. 3). However, in our case, where the depth image is often called 2.5-dimensional, we detect the boundary from the processed image with a Canny filter and acquire a two-dimensional rectangular boundary (upper image in Fig. 4). We determine the distance from the user (i.e., the depth sensor) to the object as the minimum depth value in the two-dimensional rectangular boundary (i.e., d_1 , d_2 , and d_3 in Fig. 4). We have found that some objects would result in the construction of multiple boundary rectangles (see Fig. 5). To address this problem, we apply an algorithm to aggregate smaller rectangles into larger ones if the distance values are within a reasonable range. Considering the relative positions of the bounding rectangles on each frame, we can estimate the velocity (i.e., movement direction and speed) of each object (Fig. 6).



2.4 Audio Warning and Information

By tracing each object in its respective bounding box, we can estimate the angle (θ in Fig. 7) between the user and the object from the vectors of the moving objects and the user. Based on this information, we can determine the degree of risk for each object. For example, if the distance of an object is within one meter and the angle θ is from 90° to 180° , we can conclude that the object imposes an immediate danger to the user, because it is very close and moving toward the user. However, if another object is moving toward the user with a similar angle but the distance is farther than one meter, then we can consider that this object may cause possible harm to the user, but we have more time to avoid the object. As shown in Fig. 8, our voice warning scheme provides four audio cues based on the current information from this observation. After determining the current situation, audio is provided as follows:

“[Voice information listed in Table 3 according to the situation], there is an obstacle [distance] ahead on the [“right” or “left”] side.”

For example, if an object is far from the user at 10 cm (i.e., within 1 m) on the right, and is moving away from the user at 30°, then the device warns,

“Be careful, there is an obstacle 10 centimeters ahead on the right side.”

Based on this voice information, the user can determine her/his direction to walk safely.

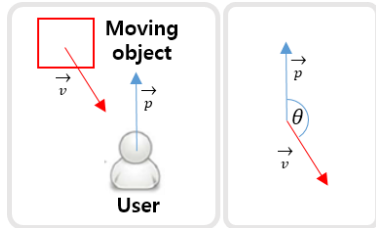


Figure 7. Relative direction between the moving object and the user

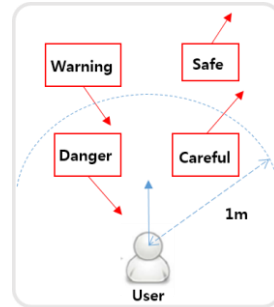


Figure 8. Criteria for degree of risk [6]

Table 3. Audio information according to the situation

| | Distance ≤ 1 m | Distance ≥ 1 m |
|---------------------------------|----------------|-----------------|
| $-90^\circ < \theta < 90^\circ$ | “Be careful” | No audio (safe) |
| $90^\circ < \theta < 180^\circ$ | “Danger” | “Warning” |

3 Results and Discussion

Figure 9 illustrates the recognized objects and their bounding boxes. In addition, Fig. 10 shows our prototype used for the proof of concept, which the user wears on her/his chest. We expect the size of the proposed device could be made much smaller once smaller depth sensors are on the market [11]. In the next step, we plan to improve our system by performing usability and usefulness tests for those who have weak vision or visual impairment.

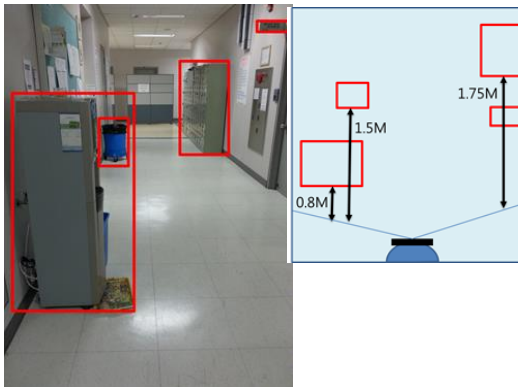


Figure 9. Objects recognized [6]



Figure 10. Prototype device [6]

Acknowledgements. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (NRF-2013R1A1A2064147)

References

- [1] Sandra Mau, N. Melchior, Maxim Makatchev, Aaron Steinfeld, *BlindAid: An Electronic Travel Aid for the Blind*, CMU-RI-TR-07-39, The Robotics Institute, Carnegie Mellon University, 2008.
- [2] Laehyun Kim, Sehyung Park, Sooyong Lee, Hyunchul Cho and Sungdo Ha, Improvement of An Electronic Aid for the Blind using Ultrasonic and Acceleration Sensors, *Journal of Korea Information Science Society: Software and Applications*, **36** (2009), no. 4, 291-297.
- [3] Kyoung-ho Kim and Sang-Woong Lee, Positioning System for the Blind Navigation, *Journal of Korean Institute of Next Generation Computing*, **8** (2012), no. 4, 6-16.
- [4] Oxford smart glasses. <http://www.ndcn.ox.ac.uk/research/oxford-smart-specs-research-group>
- [5] Kinecthesia. <http://www.kinecthesia.com>
- [6] Jee-Eun Lee, Yujin Kim and Taejung Park, Travelling Aids for Visual Impaired based on Depth Sensor, *Advanced Science and Technology Letters*, **125** (2016), 45-49. <http://dx.doi.org/10.14257/astl.2016.125.09>
- [7] OpenNI SDK. <http://openni.ru/openni-sdk>
- [8] Raspberry Pi. <https://www.raspberrypi.org>
- [9] Canny edge detector. https://en.wikipedia.org/wiki/Canny_edge_detector
- [10] Hairol Nizam Mohd Shah, Mohd Zamzuri Ab Rashid and Tam You Tam, Develop and Implementation of Autonomous Vision Based Mobile Robot Following Human, *International Journal of Advanced Science and Technology*, **51** (2013), 81-92.
- [11] Intel RealSense Camera. <http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-depth-camera.html>

Received: April 15, 2016; Published: August 11, 2016