KinDectect: Kinect Detecting Objects

Atif Khan, Febin Moideen, Juan Lopez, Wai L. Khoo, and Zhigang Zhu

Department of Computer Science, City College of New York, New York, NY 10031
{akhan14,fmoidee00,jlopez14}@ccny.cuny.edu,
{khoo,zhu}@cs.ccny.cuny.edu

Abstract. Detecting humans and objects in images has been a very challenging problem due to variation in illumination, pose, clothing, background and other complexities. Depth information is an important cue when humans recognize objects and other humans. In this work we utilize the depth information that a Kinect sensor - Xtion Pro Live provides to detect humans and obstacles in real time for a blind or visually impaired user. The system runs in two modes. For the first mode, we focus on how to track and/or detect multiple humans and moving objects and transduce the information to the user. For the second mode, we present a novel approach on how to avoid obstacles for safe navigation for a blind or visually-impaired user in an indoor environment. In addition, we present a user study with some blind-folded users to measure the efficiency and robustness of our algorithms and approaches.

1 Introduction

According to American Foundation for the Blind, the number of legally blind in the US is 1.3 million and the total number of blind and visually impaired is 10 million (100,000 students), and these numbers are increasing with the aging population in US and around the world. If we can accomplish a reliable and robust vision solution for blind people with cutting edge technology, at an affordable cost, then it will have a tremendous impact. The goal of this project is to develop a wearable computerized system that assists blind and visually impaired individuals in detecting multiple humans, and in detecting and avoiding obstacles.

Blind and visually impaired individuals encounter many challenges. One of the most common challenges is the inability to detect obstacles along their walking path. Kinect can be and has been used as a tool to help the blind and visually impaired people to detect humans and objects along their paths [1]. Xtion Pro Live by Asus is a sensor capable of acquiring both RGB color images and depth images in real-time. By combining the RGB camera and depth sensor, Xtion Pro Live offers a whole array of capabilities including: motion capture, object recognition and detection, facial recognition, voice recognition, 3D mapping, and others features which are fundamental and required to achieve our project idea. Furthermore, it can be run on USB power, which results in an ideal sensor for a wearable system.

The above features can be interconnected together to develop a robust visual guidance system which could help the blind on their daily tasks. For instance, walking by

K. Miesenberger et al. (Eds.): ICCHP 2012, Part II, LNCS 7383, pp. 588–595, 2012.

a corridor, detecting and avoiding obstacles, recognizing a human or object, finding a particular path, and many others, without the use of a walking dog (extremely expensive) or the white cane. In this paper, we utilize the depth information that Xtion Pro Live provides to detect humans and obstacles on the path of a blind or visually impaired person.

The paper is organized as the following. Section 2 discusses related work on various methods on obstacle and object detection for the blind. In Section 3, we discuss two algorithms using the 3D sensor: human detection and obstacle avoidance. In the first part, we focus on how to track and/or detect multiple humans and moving objects and then convey the information to the user. In the second part, we present a novel approach on how to avoid obstacles or objects for safe navigation of the blind or visually-impaired user in an indoor environment. In Section 4, we present a user study with some blind-folded and blind users to measure the efficiency and robustness of our algorithms and approaches. Finally, Section 5 concludes our work.

2 Related Work

There are a number of groups who have used various 3D vision techniques to solve the problem of object detections to guide blind people. Zöllner and Huber [1] designed a vibrotactile waist belt and markers from the AR-Toolkit by leveraging the Microsoft Kinect camera. Tyflos [2] is a prototype device consisting of two tiny cameras, a microphone, and ear speaker, mounted into a pair of dark glasses and connected into a portable PC for blind individuals. The overall idea is to detect changes in a 3D space by fusing range and image data captured by the cameras and creating the 3-D representation of the surrounding space.

Meers and Ward [3] proposed an electro-tactile stimulus approach with stereo video cameras and GPS for providing the user with useful 3D perception of the environment and landmarks without using the eyes. Authors in [4] used Kinect sensor on a wheeled indoor service robot for elderly assistance. The robot makes use of a metric map of the environment's walls, and manipulates the depth information of the Kinect camera to detect the walls and localize itself in the environment.

Dakopoulos and Bourbakis [5] aim to achieve obstacle avoidance and navigation in outdoor environments with the aid of visual sensors, GPS, and electrotactile simulation. Chen and Aggarwal [6] propose a segmentation scheme using Kinect to separate the human from his or her surroundings, and extract the whole contours of the figure based on the detection point.

Drishti [7] employs a precise position measurement system, a wireless connection, a wearable computer, and a vocal communication interface to guide blind users and help them travel in familiar and unfamiliar environments independently and safely. The system in [8] detects the nearest obstacle via a stereoscopic sonar system, and sends back vibro-tactile feedback to inform users about their localization at increasing the mobility of visually impaired people by offering new sensing abilities.

3 Human and Obstacle Detection

The Xtion Pro Live is built for gaming, so in most of the applications, the sensor is stationary. In our experiments, up to three people can be easily identified using the OpenNI framework, if the sensor is stationary. However, if the sensor is in motion while the user is walking, it is difficult to use the OpenNI's existing functions to detect and track people. It would be an interesting research topic to identify and track people when the sensor is in motion. In this paper we assume that the system will keep stationary when the user wants to identify people, and when the user walks, we switch to a different mode - obstacle detection and avoidance.

3.1 Human Detection and Transducing

We use the OpenNI framework [9] of the Xtion Pro Live to detect humans. We were able to identify the number of people and their relative positions in real-time from the Xtion Pro Live in millimeters (mm). For system debugging, we integrated and configured the OpenNI framework with OpenCV, an open source computer vision library [10].The program has the following key components:

1. Generate and show depth and RGB information using the OpenNI framework, which provides an application programming interface (API) for writing applications utilizing natural interaction. It covers communication with vision and audio sensors, as well as high-level middleware solutions (e.g. for visual tracking using computer vision).
2. Tracking people, and estimate distances and positions in 3D space. For tracking and estimation we take the depth value and display on the user torso.
3. People's depth data labeling and coloring for debugging.
4. Making OpenNI data (depth & RGB) accessible and compatible with OpenCV (RGB to BGR conversion)

Figure 1 shows one of the detection results, where three persons are detected. For showing that our algorithm can separate each person from the background, we use a unique color for each person that the algorithm detects.

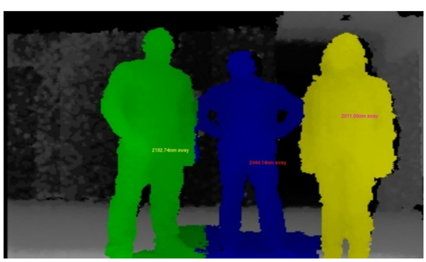

Fig. 1. Depth data: people tracking. Each colored region represents the image of a person, and the 3D location of the centroid of each person is annotated on the region of each person.

Based on this result, our system can notify the blind user about the number of people which happens to be in his or her field of view, and their respective distances and positions (in millimeters, mm). Therefore, the user gains knowledge about people in front of him/her. This information can also be conveyed to user using text to speech function via headphone [11].

3.2 Obstacle Detection and Avoidance

The obstacle detection algorithm will work mainly on the depth component of RGB-D data. In the future, we plan to fuse the depth information with the color information to improve the system's performance in obstacle detection. To improve the efficiency, we divide the original range array of 640 (H) x 480 (V) into blocks of 32 (H) x 40(V) pixels (Fig. 2). Thus, each block consists of 1280 pixels resulting in a total of 20x12 blocks of entire field of view, and we calculate an average depth within each block. Then, we further group the 20x12 blocks into 5x3 regions with each region containing 4x4 blocks. Horizontally, we have 5 directions: middle, left, far left, right, far right, and vertically we have 3: top, middle, bottom. In each region, the average of the 16 blocks will be used to give a metric that an obstacle in the region should be avoided. In our experiment, we sort the 16 average depths and calculate the average (Z) of the middle 10 values. Then, the metric of obstacle is calculated as $M = Z/Z_m$, where Z_m is the minimum depth that an object can be to the user; in our experiments, we set Z_m = 1.0 meter. After we calculate the metric of obstacle for each region, then in each direction, the largest of the three metric values are assigned as the metric of obstacle in that direction. We note this metric as M_d, d = (-2, -1, 0, 1, 2), representing far left, left, middle, right, far right.

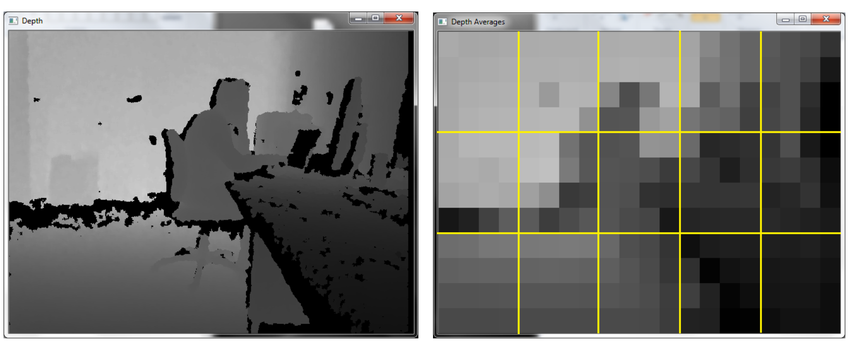

Fig. 2. Obstacle detection in 5x3 regions

We use two approaches in prompting the blind user about obstacles. In the first naive approach, we use the direction with the smallest probability of obstacle, and inform the user to turn to that direction if it is different from the previous one. This is searched in the order (far-left, left, middle, right, far-right). This raises issues that the changes of directions may be too frequent, and when two directions have the same probability, the direction that is first noted is always reported as the safe direction, with far-left is the more preferred direction.

In the second approach, in order to make the path smooth, we could use a direction update approach to give the current direction based on the previous direction, which asks the user to go to the closest direction that is still safe, e.g. $>1.5Z_m = 1.5$meters. This eliminates the problem of always preferring the far-left - here we prefer the directions in the order panning out from the middle.

4 Implementation and Experimental Results

4.1 System Architecture

The diagram in Fig. 3 shows the system schematic components. The main components are the Xtion Pro Live, the Laptop, and Bluetooth. The Xtion Pro Live, being smaller and having USB power cord, is connected to the laptop for scanning purposes. The laptop processes the depth data from the real world, and performs the detection algorithms as the blind walks along his or her passage. The computer notifies the blind person of objects ahead, and provides the user with an alternative route to avoid a possible collision. The laptop notifications are text-to-speech technology, and the user is listening to them through a Bluetooth. The user can also interact with the computer by asking if there are any people around. And the system will remind the user to stand still for a while when the system is trying to figure out how many and where the persons are.

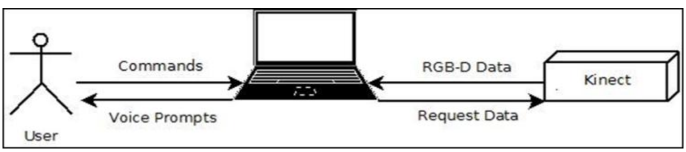

Fig. 3. System Schematic

4.2 System Integration and Testing

We installed the Xtion Pro Live on a belt for testing purposes. The prototype system includes these modules (Fig. 4), Xtion Pro live (Kinect), waist assembly to mount the Kinect, laptop for processing and transducing the data, laptop backpack to hold the laptop and headphone for giving the user directions.

We performed a user study with four blind-folded users walking in an indoor environment to validate the efficiency and robustness of our algorithms and approaches. Fig. 5 shows a picture of the scene where the users are asked to go from a start point roughly to an end point along a path. Fig. 6 shows the directions that the four users were instructed by our system using the simple one-step approach. You could see the unsmooth paths of most of the users. Fig. 7 shows the directions that the four users were instructed by our system using the direction update approach, where the paths are much smoother.

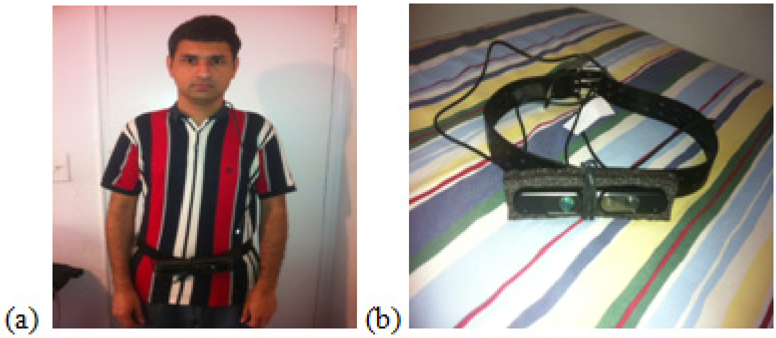

Fig. 4. Xtion Pro Live-Waist assemblies

Fig. 5. An indoor scene with the expected path labeled

The graph in Figure 6 represents obstacle avoidance paths of 4 blind-folded users using the first approach. In this diagram, the crosses represent the 4 different objects that we have: 2 trash-cans and 2 waste baskets. The X-axis shows the walking steps, each step covering 2 25cm tiles. The Y-axis demonstrates the actual number of tiles to get to that object. The coordinate (1, 12) represents the starting point; we see the viewpoint from this coordinate in Figure 5. The shifts represent the directional movements that the system prompted the blind-folded user in real time to avoid the obstacles. We observed that in this one-step approach, the algorithm favored the left-hand side first before the right-hand side. This is the reason why for some paths there is more left-turns that in others.

The plot in Figure 7 represents the experiment results of the direction update approach. In one step approach we observed that our algorithm favors the left hand side. In this plot, we observe that the middle direction is suggested to the user as long as there are no obstacles in the way. We can certainly conclude that this approach is better than one-step approach because it does not ask user to go left/right unless obstacle detected.

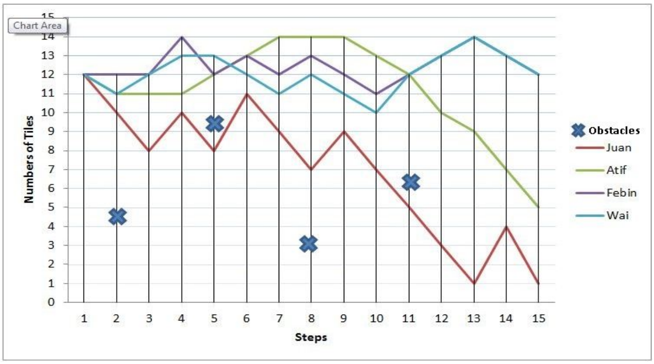

Fig. 6. Real paths of four users guided by the one-step approach

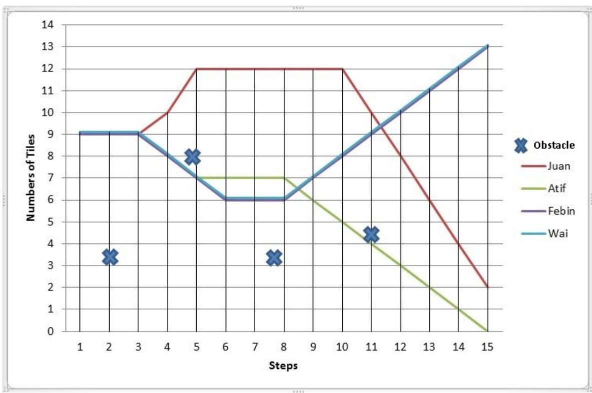

Fig. 7. Real paths of four users guided by the direction update approach

5 Conclusion and Discussions

In general, this system has the potential to replace the walking dog, which is extremely high cost, or even the white cane. As a result, this would facilitate the walking of a blind person by eliminating the fear of experiencing a collision. Compared to the current state of the art techniques that involve heavy computation to derive the depth information to achieve a simple task as obstacle avoidance, our proposed technique will be computationally simpler: Instead of building a global map, we are simply interested in a local map and we do not need a persistent map since the obstacle avoidance decision is made on the spot (i.e. locally). We also have found that the people detection function is a useful one for blind users.

However, obstacle detection and object avoidance has many limitations. For instance, most of the algorithms traditionally used for obstacle detection and navigation might work well, but when used in real time and integrated with Xtion Pro Live, they might result slow execution. Furthermore, the SAPI feature of our system reduces the computational efficiency output. Overall, obstacle detection and avoidance is still a difficult task to accomplish completely and make suitable 100% for blind users.

For people detection, the current algorithm requires the user to stand still to detect and track people, which have some constrained to the use of the sensor in real life. We hope to develop human detection algorithm when the sensor is in motion. This will create a lot of challenging issues since the sensor is in motion. All the background objects are in constant motion, thus making human detection using 3D and motion very difficult.

Acknowledgments. This paper was based on a computer science undergraduate senior design project at the City College of New York. The work is partially supported by US National Science Foundation Emerging Frontiers in Research and Innovation Program under Award No. EFRI-1137172, National Collegiate Inventors and Innovators Alliance (NCIIA) Course and Program Planning Grant (8281-10), and City SEEDs: City College 2011 Grant for Interdisciplinary Scientific Research Collaboration.

References

1. Zöllner, M., Huber, S., Jetter, H.-C., Reiterer, H.: NAVI – A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 584–587. Springer, Heidelberg (2011)
2. Bourbakis, N.: Sensing Surround 3-D Space for Navigation of the Blind. IEEE Engineering in Medicine and Biology Magazine 27(1) (January-February 2008)
3. Meers, S., Ward, K.: A Substitute Vision System for Providing 3D Perception and GPS Navigation via Electro-Tactile Stimulation. School of IT and Computer Science University of Wollongong, Wollongong, NSW, Australia (2005)
4. Cunha, J., et al.: Using a Depth Camera for Indoor Robot Localization and Navigation. DETI/IEETA- University of Aveiro, Portugal (2011)
5. Dakopoulos, D., Bourbakis, N.: Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. IEEE Trans. on Systems, Man, and Cybernetics, January 1 (2010)
6. Xia, L., Chen, C.-C., Aggarwal, J.K.: Human Detection Using Depth Information by Kinect. In: International Workshop on Human Activity Understanding from 3D Data in Conjunction with CVPR (HAU3D), Colorado Springs, CO (June 2011)
7. Ran, L., Helal, S., Moore, S.: Drishti: An Integrated Indoor/Outdoor Blind Navigation System and Service. In: PerCom, pp. 23–32 (2004)
8. Cardin, S., Thalmann, D., Vexo, F.: Wearable Obstacle Detection System for visually impaired People. Visual Computer: International Journal of Computer Graphics archive 23(2) (January 2007)
9. OpenNI Program Guide, `http://www.openni.org/Documentation/`
10. OpenCV, `http://www.opencv.itseez.com`
11. MSDNS. SAPI Function. Microsoft, August 28 (2008), `http://msdn.microsoft.com/en-us/library/aa911241.aspx`