Navigation Aid for Blind People Using Depth Information and Augmented Reality Technology

Osama Halabi  Mariam Al-Ansari  Yasmin Halwani
Fatma Al-Mesaifri  Roqaya Al-Shaabi
Qatar University, Department of Computer Science and Engineering
{ohalabi, ma095669, yh090582, fa095704, ra090350} (at) qu.edu.qa

Abstract
This paper presents a new navigation aid hybrid system for blind people based on depth information, Augmented Reality (AR), and 3D sound. The system acquires depth data in real-time from Kinect depth sensor and detect the object, then generate 3D spatial sound to convey the direction and the distance of the nearest object. The RGB camera in the same Kinect sensor is used to recognize AR markers and provide adequate information about the surrounding environment using synthesized sound. The system architecture ensures satisfactory real-time performance, which contributed to the improvement achieved towards smooth and unhindered blind travel in indoor environment, and provided adequate help information about the surrounding environment as well.

1. Introduction
We live in a world where 285 million people are visually impaired. 39 million are blind. According to World Health Organization, An estimated 19 million children are visually impaired. 1.4 million are irreversibly blind for the rest of their lives. Blindness, the most severe form of visual impairment, deprives people the ability to move about unaided [1]. Imagine walking into an unfamiliar airport. The places we have to search for, airline ticket counter, security check-in, boarding gate, are difficult to find even with signs. Imagine how much of a challenge this would be if you cannot even see the signs!

Everyday situations present similar challenges. While shopping malls often have building maps, they are usually stationary displays that are useful only when one can locate and read the display. Many medical and academic buildings lack even this kind of navigation assistance. Challenging for a sighted person, the task of finding a way in such a building for an unassisted person with visual impairment becomes nearly impossible. About 90% of the blind cannot travel independently; 7% use a white cane; 3% could use a guide dog. Though it is still very hard for them to walk around, as a perfect solution does not exist yet [2].

Many studies have been done to develop systems to assist blind people using different technology, but can be roughly classified into two categories. The first one is using robot-based guidance to detect and avoid obstacles [3]. However these guide robots limit the person’s field of activities for it cannot go through crowded streets or through irregular ground. The second approach is using intelligent cane equipped with sensors to help finding the objects [4] or using handheld infrared (IR) sensors [5]. In recent years handheld, general purpose computing devices have become increasingly affordable to the consumer. Many of these computing devices offer adequate speed, performance and resources for real time instrumentation and human-computer interface application. Maroof and his colleges utilized such a device for sensor data processing and developed a pocket-PC base navigation aid for blind individuals. The application detects surrounding obstacles using ultrasonic range sensors and generates a virtual acoustic environment where nearby obstacles become recognizable to the user. The user can perceive surrounding and direction using specialized 3D sounds [6]. However these solutions suffer from high cost for the active transmitters and receivers as well as limitation in capability and usefulness due to the fact that IR requires line-of-sight and the active transmission requires power supplies and maintenance.

In this paper we propose a new system for providing navigation aid for blind people using a new hybrid solution based on depth information, Augmented Reality (AR), and 3D sound. The system acquires depth data in real-time from Kinect depth sensor and detect the object, then generate 3D spatial sound to convey the direction and the distance of the nearest
object. The RGB camera available in the Kinect camera is used to recognize AR markers to inform the user important information about his surrounding like building name, offices, and facilities near by playing synthesized sound. Ueda et al. used a 3D scanner to acquire 3D range data map of the environment to detect the objects and gives information using synthesized sound [7]. However, it takes some seconds to acquire enough 3D range data for recognizing the environment and can only detect bumps and not walls. AR has been used to help color blindness patients who have difficulties in distinguishing object with certain colors [8, 9] but no AR-based specific application for blind people yet developed. Most of the application that utilized 3D sound based on preparing sound samples in programmed directions which then mapped to the obstacles [6]. In “Kinect for the blind” application [10], they mapped the depth image to belly-mounted tactile matrix of 32. Never the less, it does not convey the distance of the object and only gives an alert of the existence of an obstacle in front of the subject, also the feeling of the obstacle direction is questionable. Navi is another application developed by a student in University of Konstanz is similar to our work in the concept of utilizing the depth image to convey distance and AR to present information [11]. However, in Navi he uses three vibrators to convey the direction which means low resolution and limited to front, left and right feedback, meanwhile in our work we use 3D sound that can present information in all 3D direction to the user without the need for extra hardware. In addition, our system conveys not only the direction of the obstacle but also how far it is from the current location. In our system, the sound location is based on the obstacle 3D real location, which provides an accurate spatial localization feedback. The combination of these three technologies together created an efficient navigation aid solution for blind people with minimum hardware by utilizing the two cameras already available in Kinect. This means an efficient and low-cost solution.

2. Background

2.1 3D sound

Since 1907, there have been literally hundreds of technical published related to sound perception and sound localization in humans. “3D sound” can be best understood in context of localizing (or spatializing) any particular sound. As humans, we can easily determine whether or not a sound is coming from in front of us, or behind us, or from some other intermediate position around us. We can do this with amazing accuracy using two ears even if we are not facing the sound source. The term “3D sound” means to artificially create (or re-create) the auditory cues that the listener will use to determine the location of that sound in non-real environment. The primary advantage of having two ears is the ability to identify the direction of the sound. Human listeners can detect the difference between two sound sources that are placed as little as three degrees apart, about the width of a person at 10 meters [12]. In our system the 3D sound is essential component to help the user determine the location of the obstacles by generating spatial 3D sound based on the real 3D location of the obstacles that is acquired by the depth sensor. This type of feedback does not require any special training to do the mapping as it is straightforward, which is completely different from other approaches that synthesize the image using complex sound feedback that require intensive training. The vOICe is a famous example of such approach [13].

2.2 Augmented reality

Augmented Reality (AR) is a variation of Virtual Environments (VE), or Virtual Reality as it is more commonly called. VE technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. Therefore, AR supplements reality, rather than completely replacing it. Ideally, it would appear to the user that the virtual and real objects coexisted in the same space [14]. Show how the AR sample. It shows how the graphics and real objects are coexisting together. AR has been utilized in the proposed system in such a way that it recognizes the markers and play related synthesized sound to give useful guide information about the that marker’s surrounding. These markers can be used for providing visual information and graphics to sighted people as well by using mobile phones as it becomes easy to create AR application for mobile phones recently.
3. System architecture

The main functionalities of the proposed system are two functions: first one is providing 3D spatial navigation aid sound according to the distance from obstacles; the second one is providing on-site information about the surrounding location as an audible sound according to the recognized special markers. We first read the depth data from the depth sensor and analyze the data to decide the nearest object and its location and distance from the user as well. This information is then passed to the 3D sound engine to generate the 3D spatial sound according to the location of the obstacles so that the user can identify the location of the obstacle. At the same time, the sound frequency is altered according to the distant from the obstacle to convey the distance dimension to the user. The frequency will increase gradually as the user become closer to the obstacle.

The RGB camera provides a video images of the surrounding, these images are passed to the augmented reality engine to recognize special markers. Once the AR engine detects a marker it will display a 3D graphics related to that marker’s location and at the same time provide information about its location in audible sound for blind people. The system architecture can be seen in Figure 2.

![Figure 2 High-level system architecture](image)

3.1 System hardware

The system mainly needs two cameras, depth image camera to acquire the depth data and RGB camera to acquire color images of the surrounding. The Kinect camera combines the above two cameras in one small size device and in affordable low-cost price, which make it a perfect solution for our system. Kinect camera needs AC power, which makes it not suitable for portable solution; therefore we created a regulator circuit so that Kinect can operate from DC rechargeable battery. The second component of the system is headphone to enable the user to hear 3D naviga

3.2 System software

The proposed system incorporates two technologies, AR and 3D sound. Also, deals with two cameras at the same time, depth image camera and RGB camera. This complicated the implementation of the system. Many SDKs and libraries were utilized. Microsoft Kinect SDK was used to connect the Kinect camera and its hardware [15]. FMOD sound library were used to generate the 3D sound. The library offers robust functions for creating 3D sounds and it is already the default option for many game engines [16]. ARToolKitPlus was used for implementing AR functionality. ARToolKit is a software library that can be used to calculate camera position and orientation relative to physical markers in real time [17]. In addition to many APIs used to handle the camera and video as shown in Figure 4.
4. Experimental procedure and evaluation

The system has been assembled and made portable to be able to test it in real navigational tasks to evaluate the efficiency of the proposed system. We created different scenarios, each scenario represents different navigational task that might encounter the blind people in real-life situations. In each scenario we asked fully-sighted individual to wear a black thick glasses and navigate through each task. Figure 5 shows one scenario that has been tested. In this test the user is asked to detect where is the door of the class room and go towards the door and open the door, then once get inside she will navigate between chairs and white board trying to avoid these obstacles.

In the second test, the user asked to navigate in narrow passage where the angle between the door where the marker is located and the Kinect camera is very acute. This is to check the ability of the RGB camera to detect the markers in such acute angle from 3.5m distant condition as can be seen in Figure 6 (a, b). Figure 6 (A, B) show the augmented graphics on the door’s marker enlarged. It can be clearly seen that even from such distant the AR engine with RGB camera could detect the marker and provide the proper information in real-time.

Figure 4 The system software

In the first test the user were able to know where is the designated class room by getting information according to the
detected marker on the door. The 3D spatial sound feedback enabled the user getting the sense of how far the door is and walked slowly toward the door and opened it. After that while walking inside the room the 3D spatial sound enabled the user to avoid obstacles and walk through to reach the target.

In the second scenario, the detection of the marker was very accurate even with such acute angle, and again marker information along with the 3D spatial distance based feedback enabled the user to complete the task and reach the destination.

5. Discussion
The system provides a unique and innovative solution for providing blind people with navigation aid and location-based information. The system is portable and lightweight but could be even lighter if we stripped out the Kinect camera and removed the extra motor that has not been utilized in our system. This will make the system even more compact and lighter. The range of the depth camera (3.8m) was enough to cover a sufficient area for providing proper navigation aid. Usually the white cane can only give aid within 1m, but our system gives early sound feedback that make the blind people navigate comfortably within safe distant from the obstacles.

We conducted preliminary experiments to define the accuracy of the directional sound recognition and we found that it was around 40 degrees, but we think that accuracy can be increased if we tuned the location of the 3D sound. Currently, we are using the 3D location data directly from the depth image without any modification, but using some thresholds can enhance the perception of direction from the 3D spatial sound.

6. Conclusion
This paper reported the implementation of a blind navigation hybrid system that employs AR technology for detecting markers and 3D auditory display to provide 3D spatial sound based on the distance detected using depth information. The system architecture ensures portability and satisfactory real-time performance, which contributed to the improvement achieved towards smooth and unhindered blind travel in indoor environment, and provided adequate help information about the surrounding environment as well. The preliminary experiments showed promising results. We need to conduct more experiments on blind people to get valuable feedback and be able to tune the system according to their special perception. More experiments need to be done on tuning the 3D spatial sound to increase the accuracy of directional perception. In the future we can enhance the system by adding the ability to detect different objects and provide different auditory feedback about the direction of each object at the same time. This will empower the blind people with satisfactory sense of their surroundings and enable them to move more confidently in any environment.

Further more, the augmented graphics can be used to provide vision aid and color aid for visually impaired people who suffer from certain type of vision problem like color blindness.

References


