Proceedings of the 2015
IEEE Conference on Robotics and Biomimetics
Zhuhai, China, December 6-9, 2015

Assisting Blind People to Avoid Obstacles: A Wearable Obstacle Stereo Feedback System Based on 3D Detection

Bing Li, Xiaochen Zhang, J. Pablo Muñoz, Jizhong Xiao, Xuejian Rong, Yingli Tian

***Abstract*— A wearable Obstacle Stereo Feedback (OSF) System for the Blind people based on 3D space obstacle detection is presented to assist the navigation. The OSF system embedded with a depth sensor to perceive the in-front 3D spatial information in the form of point clouds. We implemented the downsampling Random Sample Consensus (RANSAC) algorithm to process the perceived point cloud, and detect the obstacles in front of the user. Finally, Head-Related Transfer Functions (HRTF) are applied to create the virtual stereo sound which represents the obstacles according to its coordinate in the 3D space. The experiment shows that OSF system can detect the obstacle in the indoor environment effectively and provides a feasible auditory perception to indicate the in-front safety zone for the blind user.**

I. Introduction

According to the World Health Organization (WHO) Fact Sheet of Visual impairment and blindness as of Auguest 2014, 285 million people are estimated to be visually impaired worldwide: 39 million are blind and 246 million have low vision [1], as shown in Fig. 1. In addition, there is an increase of the blindness statistic data in the past decade in United States, according to Blindness Statistics Data from National Eye Institute (NEI), the cases of the blindness in 2000 and 2010 in United States are around 0.9 million, and 1.3 million respectively[2]. White cane and dog guide are the simplest existing tools to help them for mobility.

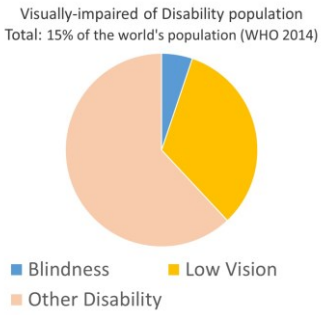

Fig. 1 Visually imparied people of the disablity wordwide

* This work was supported in part by the U.S. NSF grants CBET-1160046, IIP-1343402, and the Federal Highway Administration (FHWA) grant DTFH61-12-H-00002, and PSC-CUNY grant 65789-00-43.

Prof. Jizhong Xiao, Bing Li, Xiaochen Zhang are with the Electrical Engineering Department, CCNY Robotics Lab, The City College, The City University of New York, NY 10031 USA, (Phone: +1-212-650-7268, fax: +1-212-650-8249, E-mail: {jxiao, bli, xzhang2}@ccny.cuny.edu)

J. Pablo Muñoz is with the Computer Science Department at Graduate Center of The City University of New York, and CCNY Robotics Lab, NY 10031 USA (Email: jmunoz2@gc.cuny.edu)

Prof. Yingli Tian, Xuejian Rong are with the Electrical Engineering Department, CCNY Media Lab, The City College, The City University of New York, NY 10031 USA (E-mail: {ytian, xrong}@ccny.cuny.edu)

To help visually impaired people to live independently and improve their living quality, there are a wide range of navigation and obstacle avoidance tools available for the blind and the visual impaired. The ability of visually impaired people to access, understand, and explore unfamiliar environment will improve their inclusion and integration into the society. Among these mentioned functionalities, the obstacle avoidance is essential for the mobility of the blind users.

Electronic devices helping the Blind and the visually impaired starts from 1960s. In term of vision substitution aspect, these aids system fall into the following categories [3]:

1) Electronic travel aids (ETAs): ETAs are the devices that transform information about the environment and provide various types of feedback to the user.

2) Electronic orientation aids (EOAs): EOAs are devices that provide orientation prior to, or during the travel.

3) Position locator devices (PLDs): PLDs include technologies like GPS, European Geostationary Navigation Overlay Service (EGNOS), etc.

This paper presents and implements an audio feedback ETAs named Obstacle Stereo Feedback (OSF) system. Compared with Echolocation, Navbelt and FIU which use Ultrasonic, vOICe, Stuttgart, Virtual Acoustic Space and NAVI which use cameras, OSF system detects the position of the obstacle in front of the blind using 3D space inspection sensor, with higher obstacle inspection accuracy and reliability than the sensors like sounds, cameras, etc. Similar to FIU and Virtual Acoustic Space system, OSF system also uses head-related transfer functions (HRTF) to create a 3D stereo sound environment that represents the obstacles detected by the sensors. As to the use convenience, OSF system is a free-hands, wearable, and free-ears system which only provides reminder when detected obstacle and will not interfere the user's ability to listen to the surroundings. One of the limitation of the OSF system is that the Kinect sensor we deployed is not reliable in the outdoor environment due to the direct Sun illumination that leads to saturation in the depth acquisition. This research is based on our previous work on the indoor assistive navigation [5]-[8].

The structure of the OSF system is shown in Fig. 2. The system runs on the ROS framework, and uses Kinect to acquire 3D depth data, and then obstacle detection is processed to get the cloest points in the point cloud. Finally we generate the stereo sound based on HRTF, which is used to provide as the feedback to the user, with the direction information of the obstacle location.

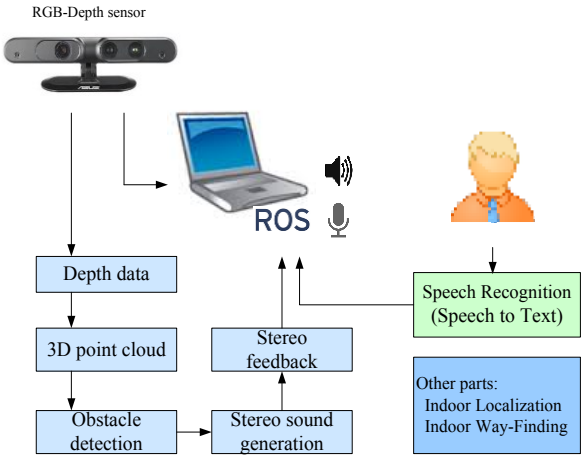

Fig. 2 Structure of the OSF in Assistive Navigation System

This paper is organized as follows: In section II, the 3D space inspection sensor and obstacle inspection techniques are elaborated. In section III, the Obstacle Stereo Feedback and HRTFs are introduced and applied to represent the obstacle position. In section V, the system implementation and experiment result of the OSF system in assistive navigation are shown. Finally, conclusion and future work are discussed.

II. Obstacle 3D Detection

In this section, we describe how to find the 3D coordinate of the obstacle from 3D depth sensor data. The position of an obstacle relative to the blind can be presented as a vector of three quantities: $p = [d, \theta_h, \theta_v]^T$, where d: obstacle distance, θ_h: obstacle horizontal direction, θ_v: obstacle vertical direction.

The position of the obstacle is determined based on information acquired by the 3D detection sensor. In this OSF prototype which is used in indoor environment. RGB-D Kinect sensor is adopted as the 3D sensor. RGB-D sensor features an RGB camera and depth sensor, and its depth sensor consists of an infrared laser projector and a monochrome CMOS sensor. Using the depth sensor, 3D point cloud data of the environment can be acquired. Next the model of the floor is detected based on the 3D point cloud data using Improved Random-down sampling RANSAC (RANdom SAmple Consensus) algorithm. Finally the obstacle is detected as the nearest off-floor point related to the blind.

A. 3D point cloud data

Online 3D perception of the front space is a crucial precondition for the reliable and safe mobility for the blind. Using RGB-D camera, we are able to present the OSF system for acquiring and processing 3D semantic information at the frame up to 30Hz which can detect obstacles, segment objects, supporting floor surfaces as well as the overall scene geometry in front of the Blind.

Depth information of the ambient space is repented as depth image in RGB-D sensor. Given a pixel $q = [u, v, d]^T$ in the depth image, where u and v are the image indices, and d is the raw depth measurement of the RGB-D camera, we can express it as a 3D point $p = [x, y, z]^T$, in the camera coordinate frame [4]:

$$z = \frac{Z_0}{1 + \frac{Z_0}{f_b} d}$$

$$x = \frac{z}{f}(u - c_x)$$

$$y = \frac{z}{f}(v - c_y)$$

Where we have the following parameters:

Z_0: distance to the reference plane used for stereo matching (an internal device parameter)

b: the baseline between the IR projector and camera

f: focal distance of the IR camera

c_x, c_y: IR image optical center

Noise of the depth sensor is handled using Gaussian mixture model by our previous research whose code is released as ROS open source package (ccny-rgbd) [13].

After we get the transferred data, which is in the camera frame. Data needs to be transferred to the world frame by rigid transformation because of sensor rotation.

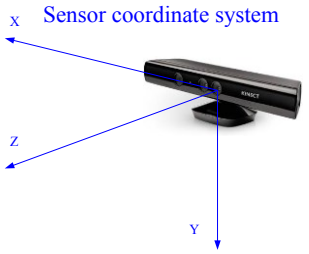

Fig. 3 Raw data frame

Given a point in camera as $p_c = [x, y, z]^T$.

In the case without considering sensor rotation, transformation matrix R_{cw}:

$$R_{cw} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

Thus point in the world frame: $p_w = R_{cw} * p$.

Next, considering the sensor rotation, the data in the world frame will be calibrated according to the Roll, Pitch, Yaw angle, which is measured by the model calculated in the next section in this paper. Some notations are defined as:

Model of the plane: $[a\ b\ c\ d]^T$, $ax + by + cz + d = 0$

Model of the floor: $model\_f = [0\ 0\ 1\ h_c]^T$, where: h_c is the height of the camera.

Rotation matrix of the floor plane:

$R_f = Rot(z, \varphi) * Rot(y, \psi) * Rot(x, \theta)$

where each Rot is rotation matrix.

$P_w = [x, y, z]^T$: a point of the floor in the before-calibration world frame.

$p_w' = [x, y, z]^T$: the coordinate of this point in world frame.

Then we have (T means transposition):

$p_w' = R_f * p_w$

$p_w'^T * model\_f = 0$

$p_w^T * model = 0$

With the calculated model_f, we can get R_f and h_c:

$$\begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} * model\_f = model$$

Finally we get the 3D point cloud data set in the calibrated world frame.

B. *Floor model segmentation*

Geometry plan segmentation is the prerequisite for obstacle detection. Once we get the 3D point cloud data, the challenge is to find the obstacle in the real-time efficiency. The first step is to remove the floor plan data. RANSAC (RANdom SAmple Consensus) is an algorithm for robust fitting of models in the presence of many data outliers, however it is highly degraded with the increase number of the data outliers. In this paper, an improved down sampling RANSAC algorithm is proposed as the solution to find the floor model effectively.

Compared with the traditional RANSAC algorithm, which is promoted by Fischler in 1981[9], and some improved RANSAC including down sampling and precertification techniques, the proposed improved down sampling RANSAC algorithm specifies the model by combining with the spatial information, which categories into down sampling approach. In this case, we assume the space position of the floor within z-max value.

Given the raw 3D point cloud data set as S, the down sampling filter is applied on each point. The down sampling data set is acquired by:

$$P = [x, y, z]^T$$

$$U \subset S$$

$$U = \{p \in S : p_z < F_{z-\max}\}$$

Where $F_{z\text{-}max}$ is set as the maximum z value of the floor.

Pseudo code for the proposed improved down sampling RANSAC algorithm:

Algorithm for improved downsampling RANSAC

1: Initialization
2: Down sampling the data set from S to U
3: **while**(steps | pre-set-cost)
4: (a) Hypothesis :
5: select randomly data subset D_k from U
6: compute model parameters $p_k = f(D_k)$
7: (b) Verification
8: compute the cost function $C_k = \sum_{x \in U} c(x, p_k)$
9: update: if $C_k < C_k^*$, $C_k^* = C_k, p_k^* = p_k$
10: **end - while**
11: Apply the model to S, to find all outliers
12: End

C. *Obstacle detection*

After applying the model to data set S to find all inlier data, obstacle position can be detected in the outlier data. In this paper, we just pick up the nearest off-floor point in front of the sensor as the obstacle for auditory feedback perception, while it is also filtered with an adjustable alarm-distance value in the horizontal plane with regard to the location of the blind.

III. OBSTACLE STEREO FEEDBACK

Human's ears can not only differentiate the sound intensity, but also predict the sound 3D direction in the environment, which involves the shadow sound which is created by the head and the reflection which is caused by the edges of the outer ears. Intuitively, we can see the influence of the outer ear edge to the input sound.

Different with monaural, and stereo sound, which are recorded by one or two microphones in empty space respectively, binaural recordings sound more realistic since they are recorded by the microphones embedded in a dummy head, in a manner which closely resembles the human's acoustic system. Thus binaural sound is closer to what human beings hear in the real word since the filter of the dummy head acts as a similar way of the human head.

To provide the obstacle position information to the blind by the 3D binaural sound, we need to synthesize accurate 3D sound. Current technology of modeling the human acoustic system have taken binaural recordings one step further by recoding the sound by the dummy head of different racial human beings. It is called head-related transfer function (HRTF) database. HRTF is a linear filter functions including a pairs (left and right) of finite impulse response (FIR) filters for specific sound positions [10].

Database of HRTFs covering the whole directions are rare except that of MIT HRTF database in US and Itakura lab HRTF database in Japan et al[11]. Here MIT HRTF database is adopted as the filter function for the spatial position information representation. MIT HRTF Database is an extensive set of HRTF measurements of a KEMAR dummy head microphone was completed in May, 1994. The

measurements consist of the left and right ear impulse responses from a Realistic Optimus Pro 7 loudspeaker mounted 1.4 meters from the KEMAR[12].

Saying we use the sound:

x(t) = the sound of "obstacle"

The pulse response of the obstacle from the HRTF database is $h_L(t)$, $h_R(t)$. Performing the signal convolution:

$$y_L(t) = x(t) * h_L(t) \quad (1)$$

$$y_R(t) = x(t) * h_R(t) \quad (2)$$

Then synthesize left and right channels are repented by:

$$y(t) = [y_L(t), h_R(t)] \quad (3)$$

IV. System Implementation and Experiment

In this section, we first illustrate the system implementation in our experiment, and then discuss the conditions and results of the experiments. In order to verify the system, a blind-fold user is participated in experiments (Fig. 4). The experiments are taken place on the ST-Hall sixth floor, Groovy School of Engineering at CCNY.

A. System implementaton

The sensors of the whole system include an ASUS Xtion PRO as RGBD Kinect senor, a Phidgets Spatial 3/3/3 as IMU and a Logitech C920 HD camera. We use a Samsung S3 laptop with speaker and microphone and a Lenovo Y510 laptop as processors. The reason of using two laptops is: the Xtion occupies 80% of the bus bandwidth such that the remaining bandwidth is unable to fulfill the demand of the rest devices. As described before, the IMU is sticking on the belt mounted RGB-D camera to get visual odometry; the camera is wearable as a hat; the laptops are placed in a backpack. Only RGBD Kinect senor depth data is used to detect obstacle.

The software is implemented under the platform of the robotics operating system (ROS) in Ubuntu, we use our previous work the ccny-rgbd-tools which is available online to perform visual odometry [13], a wrapped and simplified character appearance and structure modeling [16] to extract room numbers. We use the CMU pocketsphinx-speech-recognition [14] as the speech recognition tool, and use the sound_play in audio_common package [15] to deliver text to speech commands. The obstacle avoidance is combining with our navigation system [17].

B. Experiment

A blind-fold user wear the sensors, with laptop in the backpack are shown as below. The RGB-D sensor detects the depth info, and then convert to the 3D point cloud, as shown in Fig. 4, and Fig. 5.

Fig. 4 Experiment setup

Fig. 5 RGB-D data of the front space

Fig. 6 Down sampling data and RANSAC outcome

After the point cloud of the in-front space is acquired, we use the improved downsampling RANSAC algorithm to find

the model of the floor. The green point could are the inlier data of the model as shown in Fig. 6.

Finally the model is applied on the whole dataset as shown in Fig. 7. The green point clouds is the inlier model data, blue point clouds are the outlieres, and the red point is the position of the nearest obstacle.

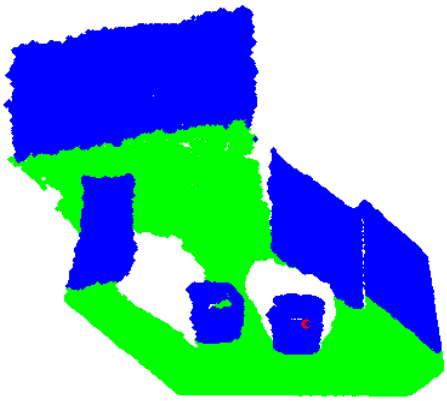

Fig. 7 Apply model and find obstacle in outliers

After acquiring the red point as the obstacle, with its coordinate in the world frame, the representation of the obstacle can be achieved: $p = [d, \theta_h, \theta_v]^T$.

With θ_h, θ_v, and corresponding HRTFs data, binaural sound for obstacle reminder is synthesized, and d is used as the sound intensity. Currently, the horizontal direction of the sound can be roughly differentiated.

The obstacle detection is intergrated in our indoor navigation system (including the acoustic feedback functionality in this research), is tested by a blind-folded user on the sixth floor of the City College Engineering Building.

V. Conclusion

In this research, we have designed and implemented a wearable Obstacle Stereo Feedback (OSF) functionality to assist indoor navigation for the visually impaired or the blind users. The experiment shows the effectiveness of the method used to detect obstacles and represent the obstacle position by auditory perception. The OSF system is capable of reliably detecting obstacles at high frame rates (30HZ). The segmentation of the floor plan is effectively to be removed to provide the obstacle position. Further research on the OSF system will be concentrated on obstacle objects recognition, such as recognizing chairs or stairs, and provide effective feedback to the user.

Acknowledgment

The authors would like to Dr. Aries Arditi from Visibility Metrics LLC for providing guidelines on this research. We acknowledge Dr. Ivan Dryanovski for his works on 3D depth sensor and visual odometry. Prof. Jizhong Xiao would like to thank the Alexander von Humboldt Foundation for providing the Humboldt Research Fellowship for Experienced Researchers to support the research on assistive navigation in Germany.

References

[1] International Agency for Prevention of Blindness. Towards universal eye health: a global action plan 2014–2019 Report. 2013.

[2] 2010 U.S. Prevalent Cases of Blindness (in thousands): Changes of Cases between 2000 and 2010. United states National Eye Institute. http://www.nei.nih.gov.

[3] D. Dakopoulos, and N. G. Bourbakis. "Wearable obstacle avoidance electronic travel aids for blind: a survey." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 40, no. 1 (2010): pp. 25-35.

[4] K. Khoshelham, and S. O. Elberink. "Accuracy and resolution of kinect depth data for indoor mapping applications." Sensors 12.2 (2012): pp. 1437-1454.

[5] S.L. Joseph, X. Zhang, D. Ivan, J. Xiao, C. Yi, Y. Tian, "Semantic Indoor Navigation with a Blind-User Oriented Augmented Reality," Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, 13-16 Oct. 2013, pp.3585-3591.

[6] S.L. Joseph, C. Yi, J. Xiao, Y. Tian, F. Yan, "Visual semantic parameterization - To enhance blind user perception for indoor navigation," Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on, 15-19 July 2013, pp.1-6.

[7] S.L. Joseph; J. Xiao, X. Zhang, B. Chawda, K. Narang, N. Rajput, S. Mehta, L.V. Subramaniam, "Being Aware of the World: Toward Using Social Media to Support the Blind With Navigation," Human-Machine Systems, IEEE Transactions on , vol.PP, no.99, pp.1,7

[8] J. Xiao, S.L. Joseph, X. Zhang, B. Li, X. Li, J. Zhang, "An Assistive Navigation Framework for the Visually Impaired," Human-Machine Systems, IEEE Transactions on , vol.PP, no.99, pp.1,6

[9] M. A. Fischler and R. C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM 24.6 (1981): pp. 381-395.

[10] C. Tonnesen, and J. Steinmetz. "3D Sound Synthesis." Encyclopedia of Virtual Environments (1993).

[11] Database of HRTFs, Suzuki & Sakamoto Lab at Japan. http://www.ais.riec.tohoku.ac.jp.

[12] B. Gardner, and K. Martin. "HRFT measurements of a kemar dummy-head microphone." (1994).

[13] I. Dryanovski, R. G. Valenti, J. Xiao. “Fast Visual Odometry and Mapping from RGB-D Data,” 2013 International Conference on Robotics and Automation (ICRA2013), May. 2013, pp. 2305-2310.

[14] http://www.speech.cs.cmu.edu/pocketsphinx

[15] http://wiki.ros.org/audio_common

[16] C. Yi and Y. Tian, “Text Extraction from Scene Images by Character Appearance and Structure Modeling,” In Computer Vision and Image Understanding, 2013, Vol. 117, No. 2, pp. 182-194.

[17] X. Zhang, B. Li, S. L. Joseph, J. Xiao, Y. Sun, Y. Tian, J.P. Munoz, C. Yi. A SLAM based Semantic Indoor Navigation System for Visually Impaired Users. IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC 2015), Accepted.