2015 IEEE International Conference on Systems, Man, and Cybernetics

# A SLAM based Semantic Indoor Navigation System for Visually Impaired Users

Xiaochen Zhang, Bing Li, Samleo L. Joseph,
Jizhong Xiao, Yi Sun and Yingli Tian
Department of Electrical Engineering
The City College of New York
New York, USA
*{xzhang2, bli, sjoseph, jxiao, ysun, ytian}@ccny.cuny.edu*

J. Pablo Muñoz, Chucai Yi
Graduate Center
The City University of New York
New York, USA
*{jmunoz2, cyi}@gc.cuny.edu*

***Abstract*—This paper proposes a novel assistive navigation system based on simultaneous localization and mapping (SLAM) and semantic path planning to help visually impaired users navigate in indoor environments. The system integrates multiple wearable sensors and feedback devices including a RGB-D sensor and an inertial measurement unit (IMU) on the waist, a head mounted camera, a microphone and an earplug/speaker. We develop a visual odometry algorithm based on RGB-D data to estimate the user's position and orientation, and refine the orientation error using the IMU. We employ the head mounted camera to recognize the door numbers and the RGB-D sensor to detect major landmarks such as corridor corners. By matching the detected landmarks against the corresponding features on the digitalized floor map, the system localizes the user, and provides verbal instruction to guide the user to the desired destination. The software modules of our system are implemented in Robotics Operating System (ROS). The prototype of the proposed assistive navigation system is evaluated by blindfolded sight persons. The field tests confirm the feasibility of the proposed algorithms and the system prototype.**

***Keywords-assistive navigation; semantic path planning; SLAM; wearable device***

## I. INTRODUCTION

According to the factsheets of the World Health Organization (WHO), 285 million people worldwide are blind or partially sighted [1]. People with normal vision orient themselves in physical space and navigate from place to place with ease. However, it is a challenging task for people who are blind or have significant visual impairment to access unfamiliar environment even with the help of electronic travel aids and vision techniques. Most of the existing travel aids transform the visual and/or range information to tactile display or audio guidance that informs the user of nearby obstacles. These devices can be cane fitted hand-held or wearable devices to warn of obstacles ahead [2]-[6] or provide 'turn by turn' guidance.

The ability of visually impaired people to access, understand, and explore unfamiliar environment will improve their inclusion and integration into the society. It will also enhance employment opportunities, foster independent living and produce economic and social self-sufficiency [7]. We realize that visually impaired people demand an assistive technology that can provide them with safe and smooth way-finding capabilities. Unfortunately, existing related assistive technologies have various drawbacks and limitations.

A number of work have been implementing inertial sensor to track and localize the users [8]-[11], however they lack accuracy and reliability for visually impaired user navigation. GPS/ GIS based approaches fit the navigation demands outdoors, but they are powerless in indoors [11]-[14].

SLAM is a process of building a map of unknown environment while at the same time localizing the robot within the map. As an extension of our previous works [15]-[18] to further improve the existing techniques in visually impaired user navigation, we propose the SLAM based wearable navigation system using multiple sensors which specifically fits the demand of visually impaired user navigation in terms of reliability. Fig. 1 shows the proposed prototype wearable system. Our system takes advantage of the SLAM technique to fuse the inputs from multiple sensors and localize the visually impaired user on the floor plan, and represents the information and guidance in a high level semantic map.

This paper is organized as follows. Section II introduces the semantic navigation system architecture and work flow; section III illustrates the detailed algorithms in the SLAM based navigation system; section IV presents the system implementation as well as the experimental results and section V concludes the paper and discusses the future research directions.

## II. SYSTEM OVERVIEW

Assistive navigations are challenging because the visually impaired user not only needs decent perception of the map of surroundings, but also demands the suitable planned path and guidance to accomplish the navigation.

Fig. 2 illustrates the system architecture of the proposed SLAM-based navigation system. The hardware includes wearable sensors (i.e., camera, IMU and RGB-D camera), interactive devices (i.e., speaker and bone headphone.) and a

This work was supported in part by the U.S. National Science Foundation under Grant CBET-1160046, the Federal Highway Administration under Grant DTFH61-12-H-00002, and PSC-CUNY under Grant 65789-00-43.

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.258

processing unit (i.e., laptop). The software is composed by nine cooperative sub-modules.


Figure 1. The proposed wearable system is composed by backpacked laptops, a microphone, a speaker, a RGB-D sensor with IMU and a head mounted camera.

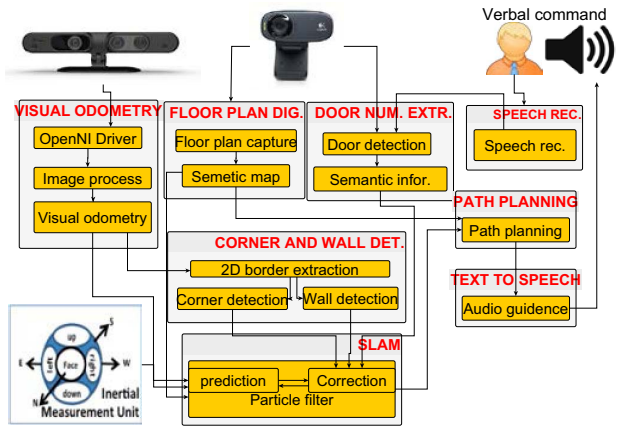


Figure 2. System architecture of the proposed frame-work.

To clarify the functionality of them, we also present the navigation scene while introducing the software sub-modules as follows. As requested by the Fire Departments in most cities in US, floor plans shall be posted at the entrances to all required exit stairs, every elevator landing, and immediately inside all public entrances to the public buildings. Thus, a visually impaired user can use the head mounted camera to scan and extract the floor plan before digitalizing it into a grid/ semantic map, using the **floor plan digitalization module,** right after he enters a building or leaves an elevator. The user needs to let the system know the destination room through the **speech recognition module**, so as to trigger the navigation. While moving, the user is localized by the **visual odometry module** using RGB-D camera [19]. The user can request the detection of room number by the **door number extraction module** using camera, when he touches the door. In the meantime, the corners will be automatically detected by the **corner and wall detection module** using depth images from RGB-D camera. The door numbers and corners are regarded as landmarks so as to match against the digitalized floor plan map as well as update the particles through **SLAM module**. The IMU is equipped to complement the orientation errors. A further orientation revising is performed through the **corner and wall detection module**. The path is planed through the **path planning module**, and be delivered to the user as motion commands and hints through the **text to speech module**.

## III. SLAM BASED LOCALIZATION

SLAM is a technology successfully used in robotic navigation, which maintains a probabilistic representation of both the subject's pose and the locations of landmarks (i.e., the "belief" of subject's location and a "map" denoted with landmarks), and refine the pose representation and map in two steps (i.e., motion and correction steps) recursively. In the motion step, the robot pose is predicted using the robot motion model. In the correction step, observations of landmarks are used to refine the probabilistic pose representation against the map, while at the same time updating the map with latest detected landmarks.

The SLAM based navigation system is built on our previous work [15]-[18]. We introduce an approach to seamlessly feed inputs from multiple sensors to the SLAM framework, so as to localize the user in the navigation scenario.

### *A. Visual odometry and local planar mapping*

We apply the previous work -- fast visual odometry using RGB-D camera [19] to provide raw pose of the user. It aligns sparse features observed in the current RGBD image against a model of previous features. The model is persistent and dynamically updated from new observations using a Kalman Filter. The algorithm is capable of closing small-scale loops in indoor environments online without any additional SLAM back-end techniques.

### *B. Visual semantics*

In order to make the user aware of their physical locations, contextual information from visual landmarks such as floor plan including signage, room number and corners are parameterized to the digitalized semantic map, as shown in Fig. 4.

#### *1) Floor map digitalization*

A heuristic method of extracting layout information from a floor plan, which employs room numbers and corners, etc., to infer landmarks and way points is used as our previous work [16]. We implement a rule-based method to localize the position of all room number labels as Fig. 5. As shown in Fig. 4, the semantic data is organized in an adjacency matrix where the coordinates indicate the connectivity between two anchors.

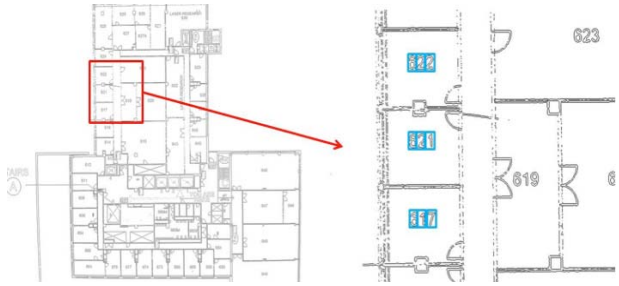


Figure. 5. The room numbers are extracted during the map digitalization [18].

*2) Landmark extraction and matching*

The visual landmarks in the immediate vicinity of the user such as room numbers and corners are extracted to localize the user.

*2.1)* we use a novel optical character recognition algorithm [20] to localize the user, when the user travels to the corresponding physical locations.

*2.2)* we use the real-time depth image from the RGB-D camera to detect corners. Specifically, we align the consequential depth images from time to time using the raw poses obtained by visual odometry. Then, the border is extracted after projecting the depth image in to horizontal plane.

### *C. Human machine interaction*

Similar to most of the peer works, we use **speech recognition** and **text to speech** to bridge the perceptions of the user and system. Details of the implementation using open source libraries [21] [22] are illustrated in section V.

### *D. Localization using particle filter*

Taking advantage of the integrated sensors, the prediction phase adopts motion estimations from visual odometry module and IMU, while the correction phase receives landmark confirmations from door number extraction module as well as corner and wall detection module.

*1) Particle filter*

Particle filter is used to estimate the pose distribution of the subject. The observation of landmarks is used to correct the particles while the odometry is used in the motion model. Samples of consequential particle filer updates are shown in Fig. 6.

*2) Floor map and landmark matching*

Notably the localization is meaningful only if it successfully localizes the user on the floor map in global map (global frame) – e.g. the digitalized floor plan in this context.

We set up a global frame on the digitalized floor map which is regarded as the ground truth. At each step, the visual odometry algorithm [19] processes the RGB-D data to estimate the pose of the visually impaired user and represent it in VO frame whose origin is located at the initial position when the system starts.

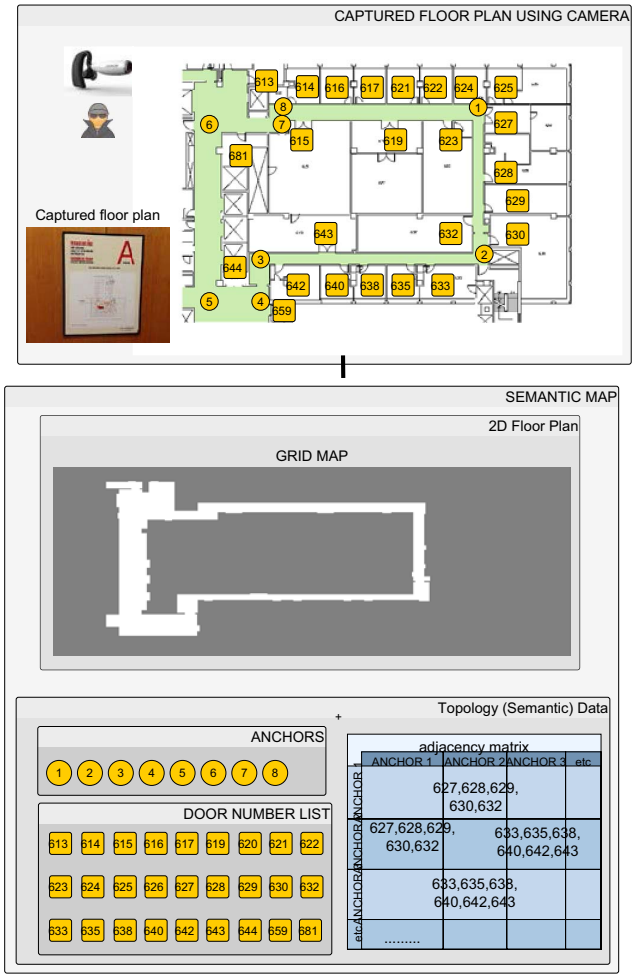


Figure 4. The semantic map is created based on a captured floor plan image. The upper half shows an image contains the floor plan is captured near an elevator, and the floor plan area is extracted. The lower half shows the digitalized semantic map and its semantic data after extraction.

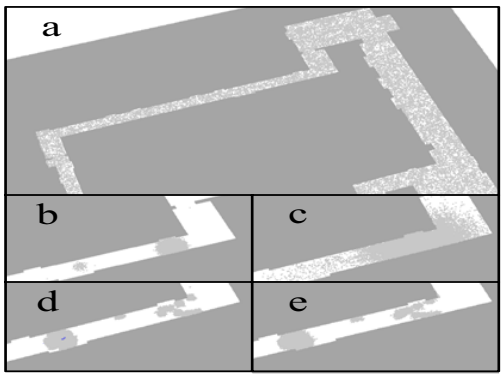


Figure 6. Samples of consequential particle filter updates on the global frame (floor plan): (a) shows the particle initialization; (b) shows the particles after the initial perception update after detecting a room by its room number; (c) shows the particles after a few steps of motion updates without knowing the heading orientation; (d) shows the particles after another perception update of another room; (e) shows the particle updates after a few steps motion updates after acknowledging the raw heading orientation.

In our work, we use room numbers and corners as the landmarks to initialize and correct the translation and rotation matrix, and further refine the rotation matrix using the depth image collected by the RGB-D camera. Consequently, the accumulated drifts are eliminated periodically.

Noting that the visual odometry poses are on the VO frame $A$; the floor map and its visual landmarks are on the global frame $W$; the IMU's poses are on the IMU frame $B$. As being widely accepted by the robotics society, we use ${}^{L}\mathbf{x}$ to denote the pose in frame $L$, composed by the planar position ${}^{L}\mathbf{v}$ and orientation ${}^{L}\theta$. Give two frames $W$ and $L$, we use ${}^{W}_{L}T$ to denote the $3\times3$ transformation matrix from $L$ to $W$, composed by a translation vector ${}^{W}_{L}t$ and a $2\times2$ rotation matrix ${}^{W}_{L}R$. Specifically, for a given rotation $\alpha$, the corresponding rotation matrix is

$$R(\alpha)=\begin{bmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{bmatrix}. \tag{3}$$

As shown in Algorithm 1, in the initial period, the system guides the user roaming around, so as to discover and detect landmarks. The user can actively trigger the room number detection of the system by verbal command after he touches a door, while the system passively detects corners.

In the "preliminary" stage, the orientation difference between the VO frame and IMU frames is recorded, as in line 3. While the user is roaming, the pose is updated accordingly, as in line 5~6. If a landmark is detected, and no prior landmark is recorded, it simply records this landmark's positions in both the VO frame and the global frame, as line 8~10. When the next landmark is detected, its corresponding positions in both frames can also be recorded. Consequently, a correspondence can be obtained to calculate the raw pose of the user, as line 12~17. The $\angle(\mathbf{v})$ in line 15 denotes the angle of vector $\mathbf{v}$; $R(\alpha)$ in line 16 denotes the rotation matrix of angle $\alpha$. After that, the stage is updated from "preliminary" to "normal". Note that, the door numbers are unique but the walls/ corners are not. Thus, in the preliminary stage, only the door numbers are accepted as landmarks. In line 21~24, the particles are updated based on new detected landmarks.

There are two potential issues causing the orientation drifts. One is that the raw user pose obtained by matching the landmarks on the VO frame and global frame is not accurate. Another is that the particle updates which revises the user pose may incurs accumulative errors. IMU can limit the accumulative drift but cannot do anything with the initial estimation error. Recall that a local planar map (Fig. 3) is kept updating while the user is moving. It is easy to obtain the angle of wall in frame $A$ by using border extraction and linear regressions. At the same time, it is feasible to find the surrounding wall's angle in the global frame $W$. Projecting the two angles onto the same frame, it is straight forward to calculate the compensation for orientation correction, as line 25~ 27 in algorithms 1. This orientation revising is not frequently triggered: on one hand, it needs to avoid the significant drift of visual odometry caused by lacking visual features; on the other hand, the accumulative orientation drift in a short period is limited since an IMU stays in the loop. Line 29~31 denotes the motion model $\delta_{trans}$ and $\delta_{rot}$ update and the corresponding particles' prediction.

Algorithm 1

```
1:  Initialization
2:  stage ← preliminary
3:  ᴬθ′ = ᴬθ − ᴮθ
4:  while (undone)
5:      updateVisualOdom(ᴬv, ᴬθ)
6:      ᴬθ″ ← ᴬθ′ + ᴮθ
7:      if (newLandmarkDetected & stage==preliminary)
8:          if (ᵂv′==null)
9:              ᵂv′ ← landmark position on W
10:             ᴬv′ ← ᴬv
11:         else
12:             ᵂv″ ← landmark position on W
13:             ᴬv″ ← ᴬv
14:             ᵂₐt ← ᵂv″ − ᴬv
15:             α ← ∠(ᵂv″ − ᵂv′) − ∠(ᴬv″ − ᴬv′)
16:             ᵂₐR ← R(α)
17:             stage ← Normal
18:         end-if
19:     end-if
20:     if (newLandmarkDetected & stage==Normal)
21:         if (door||corner)
22:             partileCorrectionPhase()
23:             ᵂₐt ← localizationStateUpdate(partiles)
24:         end-if
25:         if (wall)
26:             ᴬθ′ ← ᵂθ_wall + (ᴬθ − ᴬθ_wall) − ᴮθ − α
27:         end-if
28:     end-if
29:     (ᵂv, ᵂθ)_current = ᵂₐT(ᴬv, ᴬθ″)
30:     (δ_trans, δ_rot) ← (ᵂv, ᵂθ)_current − (ᵂv, ᵂθ)_previous
31:     partilePredictionPhase()
32: end-while
```

### *E. Path planning and audio guidance*

As long as the adjacency matrix is obtained (see sample in Fig. 5.), it is easy to draw a path to the destination, the details are not discussed here.

Note that there is no need to deliver the very detailed moving guidance to the user since he does not have to follow the optimized trajectory. Visually impaired user is prone to walk with safety, and they don't like the robot style rigid commands. The audio guidance is good enough as long as it can guide the user along a raw path towards the goal. At this

moment, the system is unable to guide the user avoiding active objects.

## IV. SYSTEM IMPLEMENTATION AND EXPERIMENTS

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 user is participated in experiments, but the data in the following analyses is based on repeated blind fold trials. The experiments are taken place on the ST-Hall sixth floor, groovy school of engineering of CCNY, as shown in Fig. 4.

### *A. System implementation*

The sensors in the system include an ASUS Xtion PRO as RGBD 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 RGBD camera; the camera is wearable as a hat; the laptops are placed in a backpack.

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 [19], a wrapped and simplified character appearance and structure modeling [23] to extract room numbers. We use the CMU pocketsphinx-speech-recognition [21] as the speech recognition tool, and use the sound_play in audio_common package [22] to deliver text to speech commands.

In the experiment, the system is able to identify the following verbal commands from user: “start”, “verify door number”, and “destination XXX” where XXX denotes a decimal number. And the system is able to deliver the following audio commands: “system ready”, “ok, go to XXX”, “you are XX meters from the target”, “hold on for detection”, “you are in front of room XXX”, “turn left at the next corner in front”, “turn right at the next corner in front”, “left turn for XXX degree”, “you reach the target” and a number of simple commands, etc.

### *B. Experiment and results*

#### *1) Localization drifts*

To quantify the drift in the localization, we have the evaluation designed as follows. An arbitrary path is given on the corridor, as the ground truth indicated on Fig. 7. The subject starts the system localization and walks along the path. The subject intentionally traverses all the landmarks on the path. Finally, the trajectory belief is compared with the ground truth.


Figure 7. An arbitrary route is designed for the test. The blue segments indicate the landmarks to be passively detected alone the path.

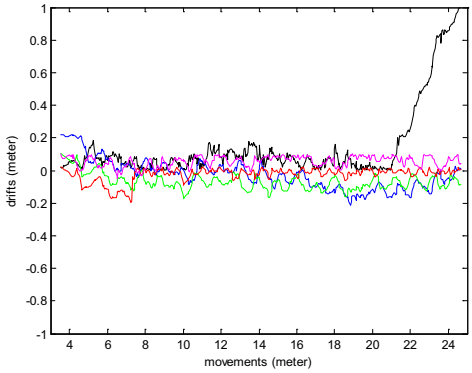


Figure 8. The localization drifts against the ground truth in five trial.

As shown in Fig. 8, the localization drifts against the ground truth is collected in five trials. It appears that the drifts are within 0.2 meters in most of the time, which is accurate enough for the navigation. Whenever a new landmark is detected on the way, the drifts can be slightly reduced. The trial in black does not converge in the figure, because the door number detection gave a wrong number on the way. If the landmarks are mis-detected, it takes a while for the particles to re-converge.

#### *2) Navigation trials*

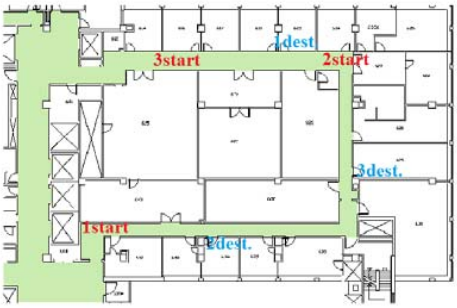


Figure. 9. The starting positions are marked in red and destinations are marked in blue. For example, the starting position and destination are noted as “1start” and “1dest.”, respectively.

The navigation system is tested by blind folded users on the sixth floor of the City College Engineering Building. Three start-and-destination pairs are given as test cases as shown in Fig. 9. In these cases, the system is capable of providing simple command and guiding the user to the destination in the simple indoor environment. The video clip of a trial can be viewed at: http://youtu.be/pj7FO41sFHM.

## V. CONCLUSION

In this research, we have designed and implemented a wearable indoor navigation aid for visually impaired users. By integrating multiple sensors of RGB-D camera, IMU, and the camera, the localization and trajectory of the user are functionally achieved using particle filter. Visual odometry from the RGB-D is corrected with the IMU odometry, and door number landmark is detected by the SVM machine learning algorithms. We also have presented a novel approach to detect the user-wall angles as the ground truth for the orientation correction, which significantly improves the fusion performance for indoor localization. Based on the localization and the semantic digital map, the user is navigated to the desired room. The user can be guided by the audio output command to the destination easily and conveniently with usability humanistic audio interface. Our future work will focus on a more effective path planning for our interactive scenario for the visually impaired people indoor navigation based on the semantic map.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Chieko Asakawa and Dr. Hironobu Takagi of IBM Accessibility Research, for providing guidelines on our application. We acknowledge Mr. Ivan Dryanovski for his works on 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] "World Health Organization – Visual impairment and blindness," (2013). [Online]. Available: http://www.who.int/mediacentre/-factsheets/fs282

[2] I. Shim and J. Yoon, "A robotic cane based on interactive technology," in *IECON 02 [Industrial Electronics Society, IEEE 2002 28th Annual Conference of the*], Nov. 2002, vol. 3, pp. 2249–2254.

[3] D. Yuan and R. Manduchi, "A tool for range sensing and environment discovery for the blind," in *Conference on Computer Vision and Pattern Recognition Workshop*, Jun. 2004, CVPRW '04, pp. 39–39.

[4] D. J. Calder, "Assistive technologies and the visually impaired: A digital ecosystem perspective," in *Proceedings of the 3rdInternational Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA'10. New York, NY, USA: ACM, 2010, pp. 1-8.

[5] R. Velazquez, E. Pissaloux, J.-C. Guinot, and F. Maingreaud, "Walking using touch: Design and preliminary prototype of a non-invasive ETA for the visually impaired," in Engineering in Medicine and Biology Society,2005. *IEEE-EMBS 2005. 27th Annual International Conference of the*, Jan. 2005, pp. 6821–6824.

[6] D. Dakopoulos and N. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C* (Applications and Reviews), Jan. 2010, vol. 40, no. 1, pp. 25–35.

[7] "World health organization (WHO): World report on disability," Jun. 2011.

[8] F. Eric, "Pedestrian tracking with shoe-mounted inertial sensors," in *Proc. IEEE Comput. Graph. Appl.*, 2005, vol. 25, pp. 38–46.

[9] A. R. Jimenez, F. Seco, C. Prieto, and J. Guevara, "A Comparison of Pedestrian Dead-Reckoning Algorithms using a low-cost MEMS IMU," in *Proc. IEEE Int. Symp. Intell. Signal Process.*, 2009, pp. 37–42.

[10] E. Z. RaulFeliz and J. G. Garcıa-Bermejo, "Pedestrian tracking using inertial sensors," *J. Phys. Agents*, 2009, vol. 3, pp. 35–42.

[11] J. Cheng, L. Yang, Y. Li, W. Zhang, "Seamless outdoor/indoor navigation with WIFI/GPS aided low cost Inertial Navigation System," *Physical Communication*, Available online 9 January 2014.

[12] J. Xiao, K. Ramdath, M. Losilevish, D. Sigh, A. Tsakas, "A low cost outdoor assistive navigation system for blind people" *Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on* , 19-21 June 2013, pp.828,833.

[13] E. Mattheiss, and E. Krajnc, "Route Descriptions in Advance and Turn-by-Turn Instructions-Usability Evaluation of a Navigational System for Visually Impaired and Blind People in Public Transport." *Human Factors in Computing and Informatics*. Springer Berlin Heidelberg, 2013, pp. 284-295.

[14] W. Jeff, Walker, B. N. Lindsay, J. Cambias, F. Dellaert, "Swan: System for wearable audio navigation," In *Wearable Computers, 2007 11th IEEE International Symposium on*, October 2007, pp. 91-98.

[15] 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.

[16] 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.

[17] Joseph, S.L.; Xiao, J.; Zhang, X.; Chawda, B.; Narang, K.; Rajput, N.; Mehta, S.; Subramaniam, L.V., "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

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

[19] 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.

[20] Y. Tian, Y. Yang, C. Yi, A. Arditi, "Toward a computer vision-based way finding aid for blind persons to access unfamiliar indoor environments," *Machine vision and applications*, 2013, *24*(3), pp.521-535.

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

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

[23] 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.