A Wearable Indoor Navigation System with Context Based Decision Making For Visually Impaired

Xiaochen Zhang^{1*}, Jizhong Xiao¹, Bing Li¹, Pablo Muñoz², Samleo L Joseph¹, Yi Sun¹, Chucai Yi² and Yingli Tian¹

¹Department of Electrical Engineering, the City College of City University of New York, New York, NY 10031, USA

²Department of Computer Science, the Graduate Center of City University of New York, New York, NY 10016, USA

Received: October 31, 2016; Accepted: November 05, 2016; Published: November 15, 2016

***Corresponding author:** Xiaochen Zhang, Department of Electrical Engineering, the City College of City University of New York, New York, NY 10031, USA, E-mail: xzhang2@ccny.cuny.edu

Abstract

This paper presents a wearable indoor navigation system that helps visually impaired user to perform indoor navigation. The system takes advantage of the Simultaneous Localization and Mapping (SLAM) and semantic path planning to accomplish localization and navigation tasks while collaborating with the visually impaired user. It integrates multiple sensors and feedback devices as an RGB-D camera, an IMU and a web camera; and it applies the RGB-D based visual odometry algorithm to estimate the user's location and orientation, and uses the IMU to refine the orientation error. Major landmarks such as room numbers and corridor corners are detected by the web camera and RGB-D camera, and matched to the digitalized floor map so as to localize the user. The path and motion guidance are generated afterwards to guide the user to a desired destination. To improve the fitting between the rigid commands and optimal machine decisions for human beings, we propose a context based decision making mechanism on path planning that resolves users' confusions caused by incorrect observations. The software modules are implemented in Robotics Operating System (ROS) and the navigation system are tested with blindfolded sight persons. The field experiments confirm the feasibility of the system prototype and the proposed mechanism.

Keywords: Assistive Navigation; Context Based Decision Making; Wearable System; Visually Impaired;

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 them 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].

After consulting with the blind individuals in organizations such as the Lighthouse International NY, Computer Center for Visually Impaired People (CCVIP) at CUNY Baruch College, New York Institute for Special Education (NYISE), and the New York State Commission for the Blind and Visually Handicapped (CBVH), 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.

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 [17-20, 37] to further improve the existing techniques in visually impaired user navigation, we present the SLAM based navigation system using multiple sensors which specifically fits the demand of visually impaired user navigation in terms of reliability. Figure 1 shows the system at work. It takes advantage of the SLAM technique to fuse the inputs from multiple sensors and localize the user on the floor plan, and represents the information and guidance in a high level semantic map which contains necessary abstracted information for human being.

Path planning for visually impaired is different with those for robots. There are numerical optimal/suboptimal solutions that handle uncertainty and incomplete information for robots. However, as a human-centered system, the navigation system shall seek to provide smooth and reliable user experience while considering the facts that visually impaired users are prone to be confused if conflicted or rigid commands are presented. Thus, we propose a context based decision making mechanism in path planning which fits the need of the navigation system. It takes advantage of the fact that the goal is to reach the destination

Figure 1: The system at work (a) A team member wearing the system (b) The RGBD sensor with slot which can stick onto the belt. An IMU is mounted on the top left of the RGBD sensor (c) The camera on the head.

while it is not necessary to guarantee the perceptions are always correct during the trip. The mechanism is flexible and capable of being tailored to specific cases based on needs. This work is an extension of our previous works [28-30, 37].

Contributions of this work are two folds. First, we developed a wearable navigation system that guides the user navigating from place to place. To the best of our knowledge, this is the first system that performs SLAM and semantic path guiding on the captured floor map. Second, we proposed a context-based path planning mechanism that efficiently utilizes the user-oriented data with uncertainty and specifically fits the need of visually impaired navigation. The path planning mechanism takes advantage of the property that motions can always be correct even if the partial perceptions during the trip are wrong. To the best of our knowledge, this is the first path planning mechanism that provides smooth and reliable guidance while wrong observation on landmarks may be presented in the context of visually impaired navigation. The prototype is implemented and tested.

This paper is organized as follows. related works; system architecture and illustrates the details and algorithms in the SLAM based navigation; context based decision making on path planning; system implementation with experimental results and ends with conclusion and the future works.

Starting from section III, the term “visually impaired user” with “user” and “door number” with “room number” are interchangeably used.

Related Works

A number of works implemented inertial sensor to track and localize the users [9-11]; however they lack accuracy and reliability for blind user navigation. GPS/ GIS based approaches fit to outdoor navigation demands but are powerless in indoor applications [12-15].

Yang et al. proposed a localization system that uses

accelerometer embedded in a mobile phone to track a person in a pre-collected RSS (received signal strength) map [16]. However, the RSS finger print based corrections requires considerable pre-installations and pre-processing in creating the RSS map. Moreover, it uses step counters and compasses for pose estimation, which has accumulative errors over time. The same problem happens to the localization and tracking system using WiFi signal proposed by Kannan [17].

In recent years, there have been a few attempts to apply SLAM technology in assistive navigation for visually impaired people. The SWAN project [15] developed a software engine using the Monte Carlo Localization and count on GPS for pedestrian outdoor navigation. The research emphasis is placed on sonification to guide the blind user; leaving many research issues untouched such as how to detect environment features for updating believe.

The “Navatar project” [18] employs an IMU in the smartphone and use a particle filter algorithm to help visually impaired persons for indoor navigation. However, it assumes the indoor map is known as a priori through manual annotation and uses human as sensor to detect simple landmarks (i.e., intersections of hall ways and open doors). This landmark confirmation by the user is prone to false alarms since the multiple landmarks may not be distinguishable by touch.

Lan et al., proposed an indoor localization system that tracks a person using the accelerometer embedded in a waist mounted mobile phone. It utilizes a floor-plan map to calibrate the direction errors from the gyroscope by comparing the layout of the building and the trajectory taken by the user [19]. However, the localization is based on dead reckoning and lacks landmark correction.

A Co-robot Cane prototype is developed to aid the visually impaired navigation. It uses a mounted 3-D camera to estimate the pose, and a web-camera to recognize the major objects while traveling. A reasonable possibility of success is proved in the experiment. However, how to handle to failures is not mentioned.

Researchers have made significant progress in applying advanced technologies in computer vision, robotics and artificial intelligence to improve mobility and navigation of people with special needs Y.H. Lee et al [20-22]. used a Kinect sensor to extract orientation information of the blind users, by incorporating visual odometry and feature based metric-topological SLAM [27]. An improved work utilizes the spatial feature and IMU to guarantee the correctness of localization, under the condition that the scenario has to be a traveled place [38].

Although a good number of navigation solutions for visually impaired are presented, little of them provides a feasible solution while incorrect observations are presented. The interactions are not truly considered human-centric since the needs of visually impaired are not taken good care. Be different with normal sighted user, the visually impaired may not be able to provide that feedback to correct error caused by sensor. On the other hand, incorrect observation may frustrate the user while navigating, especially when being guided walk back and forth. Fortunately,

in indoor navigation, the paths to the destination are very limited, one or more wrong perceptions on landmarks do not instantly lead to a different path. As long as later perceptions can correct the overall understanding of the on environment so that the user can be guided to the destination smoothly.

Our system uses a web camera and a 3-D camera to percept the landmarks to localize the user on a digitalized floor plan. A semantic map is abstracted to make the navigation feasible. A context based path planner generates path and motion guidance to user.

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 refines recursively the pose representation and map in two steps (i.e., motion and correction steps) [23-26]. 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 on the map while simultaneously updating the map with latest detected landmarks.

The SLAM principle is analogous to the navigation scenario where a visually impaired user is required to find a conference room in an office building. If a floor plan is accessible, i.e., captured by head mounted camera or downloaded from the internet, the system can digitalize it into a map before performing path planning and navigation. The system can localize the walking user by comparing the salient landmarks (e.g., door numbers) on the map. The motion and observation steps continue until the user reaches the target place.

System Overview

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

Figure 2 illustrates the system architecture. The hardware includes wearable sensors including camera, IMU and RGBD camera, interactive devices including speaker and bone headphone and processing units including a laptop. The software is composed of nine cooperative sub-modules. As follows, to clarify their functionality, we introduce the software sub-modules in the presence of a navigation scene.

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, the user can use the head mounted camera to scan and extract the floor plan before digitalizing it into a grid/semantic map, by 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 [21]. The user

Figure 2: The system architecture.

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 with 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 planned through the **path planning module**, where the context based decisions are made and delivered to the user as motion commands and hints through the **text to speech module**.

Visual odometry and local planar mapping

We apply our previous work -- fast visual odometry using RGB-D camera [8] to provide raw pose of the user. It aligns sparse features observed in the current RGB-D image on a model of previous features. The model is persistent and is 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; but like all other visual odometry approaches, it has accumulative drifts over long time in the motion from place to place. The drifts are revised using the IMU and further corrected using the wall border extracted from depth images as described later.

At the same time, the depth images which are obtained by the RGB-D sensor and presented in the form of 3D point clouds are projected onto the planar coordinate and then stacked together based on the corresponding poses estimated by the visual odometry. Particularly, the system keeps only the depth planar map with respect to the recent poses within certain travel displacement as shown in Figure 3.

Visual semantics

In order to make the user aware of their physical locations, contextual information from visual landmarks such as floor plan with signage, room number and corners are parameterized to the

Figure 3: Recent planar depth images are reorganized in a planar depth local map according to user's poses in visual odometry.

digitalized semantic map as in Figure 4.

Floor map digitalization: A heuristic method of extracting layout information from a floor plan that employs room numbers and corners, etc. to infer landmarks and way points are used [29]. A rule-based method is implemented to localize the position of all room number labels as in Figure 5. Then the range of the rooms and positions of their doors in the floor map are searched by using a vertical and horizontal scan from the region of room number. Assuming that all the rooms have their doors on the hallway, anchor points are generated by using the room number labels and corners.

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.

A) An optical character recognition algorithm [31] is used to localize the user when the user travels to the corresponding physical locations. Specifically, the door number is extracted after the user actively triggers the recognition through a verbal command "door number detection" after he perceives the existence of the door through the sense of touch. The benefits are of two folds: one is to minimize the false alarm of unwanted misdetection; another is to greatly save the processing power by image processing. If the detection fails to get any meaningful result, it will prompt the user to trigger detection again.

B) The real-time depth images from the RGB-D camera are used to detect corners. Specifically, the consequential depth images from time to time are aligned using the raw poses obtained by visual odometry. Thus, at a given time, the depth image to be processed is the stacked depth image after overlapping recent depth images according to their poses in the visual odometry. The period is restricted by the measurement of physical displacement calculated through visual odometry. Since the visual odometry drift in a short period is not heavy, the resulting stacked depth

image is a good approximation to the shape of the surrounding as shown in Figure 3.

Then the border is extracted after projecting the depth image into horizontal plane. A revised Hough Transform [32] after the Harris operator [33] is applied to obtain the corner candidates from the image of extracted borders. The corner candidates are confirmed as detected landmarks only if being

Figure 4: The semantic map is created based on a captured floor plan image. The upper half shows an image containing the floor plan captured near an elevator. The lower half shows the digitalized semantic map and its semantic data in a form of adjacency matrix.

Figure 5: Room number extraction in floor plan digitalization.

further confirmed by a shape filter of the stacked depth image. The intuition is straight forward: since the stacked depth images are always within an arbitrarily given displacement, it is easy to perceive corners because the depth image is changing from long and thin to short and thick (or long and thick) when approaching a corner.

After being confirmed, the corner candidates are matched with the corner sets in the floor map to update the particles. Particularly, the matching criterion is loosened, and thus a corner perception will enlarge the weight of particles around a few similar corners. This eliminates the possible failure caused by false alarm and yet is sufficient in application.

C) At the same time, the scale between the metrics in the floor plan and the real time scan can be obtained. As shown in Fig. 6, the right hand side is the planar surrounding obtained from the depth image of RGB-D while detecting a landmark and the left hand side is the corresponding surrounding of the same landmark on the floor plan. By comparing the width of the corridor next to the corresponding landmark, the scale between the metrics in the real time navigation and the floor plan can be easily calculated and recursively updated.

Human machine interaction

Speech recognition module and text to speech module are used to bridge the perceptions of the user and the system. Details of the implementation using open source libraries [34, 35] are illustrated in Implementation and Experiments section.

Localization using particle filter

If the digitalized floor plan is treated as a map and the doors and corners as landmarks, the localization of the user on the floor plan with doors and corners labels is analogous to the localization of a moving object on the map with preregistered landmarks. 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 as well as corner and wall detection.

Particle filter: Particle filter is used to estimate the pose distribution of the subject. x is used to denote the state, u the motion input from visual odometry, s the measurement of door number, z the measurement of corner and λ the measurement of wall angle.

The localization state is represented through a set of particles in the 2D planar space $\{x^{(i)}=(x, y, \theta)^{(i)}\}_{i \in \mathbb{N}^2 \times [0...2\pi)}$ as

$$p(\mathbf{x}) \approx \frac{1}{N} \sum_{i=0}^N \delta_{\mathbf{x}^{(i)}}(\mathbf{x}) \quad (1)$$

Where (x, y) represents the position, θ is the orientation angle of each particle and $\delta_{\mathbf{x}^{(i)}}(\cdot)$ is the impulse function centered at the particle $\mathbf{x}^{(i)}$. The denser the particles in a state space region, the higher the probability that the subject is in that region.

The distribution is represented through a set of weighted particles $\langle w^{(i)}, \mathbf{x}^{(i)} \rangle$. The particles are drawn from a proposal distribution $q(\mathbf{x}_t | z_{0:t}, s_{0:t}, \lambda_{0:t}, u_{0:t})$ by posterior distribution $p(\mathbf{x}_t^{(i)} | z_{0:t}, s_{0:t}, \lambda_{0:t}, u_{0:t})$, while the weights for each particle are computed according to the Importance Sampling Principle (ISP):

$$w_t^{(i)} = \frac{p(\mathbf{x}_t^{(i)} | z_{0:t}, s_{0:t}, \lambda_{0:t}, u_{0:t})}{q(\mathbf{x}_t^{(i)} | z_{0:t}, s_{0:t}, \lambda_{0:t}, u_{0:t})} \quad (2)$$

By choosing the motion model $p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, u_t)$ as the proposal distribution, the weight update approximately becomes

$$w_t^{(i)} \propto p(z_t, s_t, \lambda_t | \mathbf{x}_t^{(i)}) w_{t-1}^{(i)}.$$

The motion model in the prediction phase is based on the output of visual odometry, which is composed of translation model $\hat{\delta}_{\text{trans}} \sim \mathcal{N}(\delta_{\text{trans}} | \Sigma)$ and rotation model $\hat{\delta}_{\text{rot}} \sim \mathcal{U}(\delta_{\text{rot}} - a, \delta_{\text{rot}} + a)$ where $\mathcal{N}(\cdot)$ and $\mathcal{U}(\cdot)$ denote the bivariate normal distribution and uniform distribution, respectively. δ_{trans} and δ_{rot} denote the translation and rotation estimation from visual odometry, Σ and a denote the corresponding covariance matrix and restriction parameter obtained through sensor calibration, respectively. To project the visual odometry estimations in its local frame on to the global frame, a transformation after landmark matching is indispensable and to be described later.

The perception model depends on the specific sensor and task, and is assumed as a joint distribution of three independent types of perceptions, e.g. corner, room number and wall angle,

$$p(z_t, s_t, \lambda_t | \mathbf{x}_t) = p(z_t | \mathbf{x}_t) p(s_t | \mathbf{x}_t) p(\lambda_t | \mathbf{x}_t).$$

Floor map and landmark matching: The localization is meaningful only if it successfully localizes the user on the floor map.

In this 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 RGBD camera.

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 [8] process the RGB-D data to estimate the pose of the user and represent it in Visual Odometry (VO) frame whose origin is located at the initial position when the system starts. After the system detects two landmarks (e.g., doors or corridor corners), the line segment in VO frame is used to represent the actual heading of the user during last travel period. By matching it with the corresponding line segment in the floor map, the initial estimate of the user's pose can be obtained. However, the

Figure 6: Scale between the two metrics of the visual odometry frame and the floor plan frame is obtained by comparing h1 and h2.

Figure 7: Samples of consequential particle filter updates on the map (a) The particles after initialization (b) The particles after the initial perception update after detecting a room by its room number (c) The particles after a few steps of motion updates without knowing the heading orientation (d) The particles after another perception update of another room (e) The particle updates after a few steps motion updates after acknowledging the raw heading orientation.

visual odometry suffers from accumulative drifts especially in featureless environment. The IMU reading is used to correct the heading angle estimate while the user is walking. Then the particle filter is applied to refine the estimation in two phases (motion/prediction and correction/measure) iteratively. While the user is moving, its pose is predicted by using visual odometry in VO

frame and then transformed to the global frame. The particles are updated accordingly. Whenever a landmark is detected, the corresponding perception model will be used in updating the particle states.

Note 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 widely accepted by the robotics society, ${}^L\mathbf{x}$ is used to denote the pose in frame L , composed by the planar position ${}^L\mathbf{\delta}$ and orientation ${}^L\theta$. Give two frames W and L , W_LT is used to denote the 3×3 transformation matrix from L to W , composed by a translation vector ${}^W_L\mathbf{t}$ and a 2×2 rotation matrix W_LR . Specifically, for a given rotation α , the corresponding rotation matrix is

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

As shown in Algorithm 1, in the initial period the system guides the user roaming around in order to discover and detect landmarks. The user can actively trigger the room number detection of the system by a 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 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 $\angle(\mathbf{\delta})$ in line 15 denotes the angle of vector $\mathbf{\delta}$; $R(\alpha)$ in line 16 denotes the rotation matrix

Figure 8: Samples of the wall angle estimation: The green dots indicate the extracted border points; the red lines indicate the extracted line segments of wall.

of angle α . After that, the stage is updated from “preliminary” to “normal”. Note that, the door numbers are unique but the walls and 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. First, the raw user pose obtained by matching the landmarks on the VO frame and global frame is not accurate. Second, the particle updates which revise the user pose may incur accumulative errors. IMU can limit the accumulative drift but cannot do anything with the initial estimation error. Recall that a local planar map (Figure 3) is being updated while the user is moving. It is easy to obtain the angle of wall in frame A by using border extraction and linear regressions as shown in Fig. 8. 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 (Figure 9). 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 δ_{trans} and δ_{rot} updates and the corresponding particles' prediction.

Context Based Decision Making on Path

The observation result has a direct impact on the perception. Furthermore, the determination of status updates is crucial to the consequential decision making on motion. It is understandable that oversimplification of the decision making might lead to an inappropriate and even wrong decision. However, the effort on the human centered decision making has been less paid. We

Figure 9: The intuition on orientation correction is that the wall border from RGB-D data is obtainable, and the corresponding wall border on floor map is given. Aligning them corrects orientation drift.

Algorithm 1

```

1: Initialization
2: stage ← preliminary
3:  ${}^A\theta' = {}^A\theta - {}^B\theta$ 
4: while (undone)
5:   updateVisualOdom( ${}^A\mathbf{d}, {}^A\theta$ )
6:    ${}^A\theta'' \leftarrow {}^A\theta' + {}^B\theta$ 
7:   if (newLandmarkDetected & stage==preliminary)
8:     if ( ${}^W\theta' \neq \text{null}$ )
9:        ${}^W\mathbf{d}' \leftarrow$  landmark position on  $W$ 
10:       ${}^A\mathbf{d}' \leftarrow {}^A\mathbf{d}$ 
11:     else
12:        ${}^W\mathbf{d}'' \leftarrow$  landmark position on  $W$ 
13:        ${}^A\mathbf{d}'' \leftarrow {}^A\mathbf{d}$ 
14:        ${}^Wt \leftarrow {}^W\mathbf{d}'' - {}^W\mathbf{d}'$ 
15:        $\alpha \leftarrow \angle({}^W\mathbf{d}'' - {}^W\mathbf{d}') - \angle({}^A\mathbf{d}'' - {}^A\mathbf{d}')$ 
16:        ${}^A_R \leftarrow R(\alpha)$ 
17:       stage ← Normal
18:     end-if
19:   end-if
20:   if (newLandmarkDetected & stage==Normal)
21:     if (door||corner)
22:       particleCorrectionPhase()
23:        ${}^Wt \leftarrow$  localizationStateUpdate(particles)
24:     end-if
25:     if (wall)
26:        ${}^A\theta' \leftarrow {}^W\theta_{wall} + ({}^A\theta - {}^A\theta_{wall}) - {}^B\theta - \alpha$ 
27:     end-if
28:   end-if
29:    $({}^W\mathbf{d}, {}^W\theta)_{current} = {}^Wt({}^A\mathbf{d}, {}^A\theta')$ 
30:    $(\delta_{trans}, \delta_{rot}) \leftarrow ({}^W\mathbf{d}, {}^W\theta)_{current} - ({}^W\mathbf{d}, {}^W\theta)_{previous}$ 
31:   particlePredictionPhase()
32: end-while

```

suggest that the decision making should be mostly treated as a complicated problem, the scope of which should be expanded to include the user character in the loop. Without an appropriate approach to deal with the complexity of evaluation, it is likely to come out with a best decision for robots but a terrible one for the user.

Particularly in this work, it is realized that the room number detection may result in incorrect results by a number of factors: limitations on text extraction accuracy, interferes presented within the view scope (Fig. 10), etc.

An incorrect detection may lead to a temporary loss in localization, and might confuse the user if the system tries to deliver the commands with regarding to the “shortest path” to the user since it may leads to back and forth movements. For visually impaired user navigation, a smooth user experience is on the top of priorities since the user may not be able to provide as much as feedback to correct the system as normal sighted people, and is prone to feel less secure when receiving contradictory commands. Fortunately, the special feature of indoor navigation and the existing digitalized data enable a smooth and reliable decision making on path and motion planning.

If cells in the adjacent matrix (corridor between anchors) are regarded as the basic units, the navigation can be treated as the traversal game from the starting cell to the destination. It is obvious that the solution is limited. In other words, even if the current perception is incorrect, the movement may still be correct. Under the assumption that the incorrect observation is the minority, the system is guaranteed to recover the correct perception after multiple detection attempts on the trip.

Given $P(S_{ij} | O_k)$ as the probability of cell_{ij} (i-th row and j-th column) after k time's independent detections where the detection set $O_k = \{o_1, o_2, \dots, o_k\}$.

The confidence function $C(S_{ij})$ is formulated as

$$C(S_{ij}) = \sum_{i \neq j} P(S_{ij} | O_k) \cdot f(S_{ij}) - \sum_{i \neq j} b \cdot P(S_{ij} | O_k) \cdot f(S_{ji})$$

Where $f(S_{ij}) = 1 + a - d(S_{ij}) / \sum_{i \neq j} d(S_{ij})$; $d(S_{ij})$ denotes the

forward distance; a and b are empirical fraction factors which control the preference on short distance and how likely to let the user turn back. The forward distance is defined as the distance of the shortest path that does not contain cell_{ij} from the j-th anchor to the destination. In the case that no forward distance is presented, $d(S_{ij}) = \text{null}$ and $f(S_{ij}) = 0$. By comparing the $C(S_{ij})$ with a preset confidence threshold χ , the system is able to make decisions on whether to let the user move along cell_{ij}.

An intuitive explanation on how the decision maker works is as follows (see figure 4). Assume that the user walks on cell₃₂, and has some detections on landmarks. Due to some incorrect detection, the user has a major belief on cell₅₆ and minor believes on cell₃₂. Although the localization may be temporarily lost, the system may prompt the user to move forward since it leads to the anchor closer to the destination. After several detections, the localization is back to be stable and the user's previous motions are identical to the desired motions. In another case, the belief on cell₅₂ and the believes on cell₂₃ are both major. Either keep moving forward or turn backward depends on the factor b. Note that when the confidence on localization is lower than a threshold, the system will frequently warn the user to perform room number detection in order to gain more observations on landmarks.

As one of the major topics in visually impaired navigation, the audio guidance principles should follow the principles of

Figure 10: Example of interferes on room number detection.

scientific soundness, feasibility, and effectiveness so as to include indispensable functionalities related to obstacle avoidance, user preference and conciseness. The dedicated audio guidance is studied in another work focusing on human sensing and visibility metrics.

Implementation and Experiments

In this section, we first illustrate the implementation in experiment, and then discuss the results and reason. In order to verify the system, a blind user participated in experiments but the data in the following analyses is based on repeated blind fold trials. The experiments were carried out on the ST-Hall sixth floor in the Grove School of Engineering of CCNY as shown in Figure 4.

System implementation

The hardware includes an ASUS Xtion PRO as the RGB-D sensor, a Phidgets Spatial 3/3/3 as the IMU and a Logitech C920 HD camera. We use a Samsung S3 laptop with speaker and microphone as the human machine interaction base and a Lenovo Y510 laptop as a data processing center. An additional laptop is used since the bus bandwidth cannot handle all the inputs flow. As described previously, the RGB-D camera is plugged onto the user's belt, the IMU is pasted on the RGB-D camera, the web-camera is head mounted, and the laptops are backpacked.

The software is implemented on the platform of the robotics operating system (ROS), the ccny-rgbd-tools are used as the base of visual odometry [8], a wrapped character appearance and structure modeling [30] is applied to extract room numbers. The CMU pocketsphinx-speech-recognition [34] is taken as the speech recognition tool, and the text to speech is achieved using default package [35] in ROS.

Table 1: Exam Table of Room Number Detection.

	Empirical Setting	Natural Setting
First time success	0.84	0.63
Multi-time success	0.92	0.86
Failures	0.08	0.07

Figure 11: An arbitrary route is designed for the test. The blue segments indicate the landmarks to be passively detected along the path.

Experiment and results

We first perform sub-system test on landmark detection and localization, and finally perform the task oriented trail studies.

Room number detection: The first experiment is to exam the successful rate of room number detection under different conditions.

First, the camera is mounted on the head of a blind folded tester standing 60 centimeters away perpendicular to the room number, and heading towards to it. The tester is not restricted to keep steady during test. The detection is performed 20 times for each of 10 different rooms. Second, the tester moves along the corridor, turns around and triggers the room number detection 20 times for each of 10 rooms so as to exam the successful rate of room number detection in an almost realistic application. As a part of the landmark detection, the system will notify the user if the detection fails.

In Table 1, the successful rates are listed. The first time success denotes that the user gets a correct output of room number after a single trigger of detection. Multi-time success denotes that the room number can be detected after the user triggers multiple times of detection for the same door. The failure denotes a wrong detection of a room number.

Obviously, the room number detection is reliable. One of the reasons is that the outputs are restricted by the possible room number data sets obtained from floor plan digitalization. The first time success rate during the test is not very high while any motion caused by user may interfere the image quality captured by the camera thus leading to a less accurate detection. Additionally, the

room number does not always stay within the view of the camera. However, after perceiving that the room number is not detected, the user can trigger another round of detection, which enhances the chance of obtaining the true room number. On the other hand, the user may adjust his pose to present a better position. The user may leave after being noticed that the room number is detected and so the success rate plus failure rate does not equal to one.

Localization drifts: To quantify the drift in the localization, we designed an evaluation procedure as follows. An arbitrary route is given on the corridor as indicated on Figure 11. The subject starts the system localization and walks along the path. The subject intentionally traverses all the landmarks on the path. Finally, the drift in localization is calculated.

As shown in Figure 12, the localization drifts against the ground truth are collected in five trials. It appears that the drifts

Figure 12: The localization drifts.

Figure 13: The orientation drifts comparison.

Figure 14: 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.

Table 2: Average Complete Time (s).

	Original	Noise Added
CASE ONE	207	221
CASE TWO	285	332
CASE THREE	141	162

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 room number detection reports a wrong landmark near the end. It takes a while for the particles to converge.

Orientation drifts: A number of existing peer works suffer from orientation drifts which potentially make the loop closure more difficult. This work is inherently invulnerable to the orientation drift because of the fact that at any time the real time wall border obtained by RGB-D can be aligned with the wall border on the digitalized floor map as shown on Figure 8 and Figure 9.

To show this, we setup evaluations under two different conditions: one is with this orientation compensation, another is not. The test process is the same as the last test for evaluating the localization drift. The system runs two individual localizations in the background with and without the orientation compensation. The orientation drifts against the ground truth are collected, and then averaged for both cases. Figure 13 shows that applying the orientation compensation greatly remove the orientation drifts.

Navigation trials with context based decision maker: The system is tested by blind folded users on the sixth floor of the Engineering Building in CCNY. Three start-and-destination pairs are chosen as the test cases shown in Figure 14 (refer to Figure 4 for landmarks). The experiments are designed to test the ability that the system is able to guide the user to accomplish the tasks. Then, under the same experiment setup, a noise

generator is added to the room number detection module. It has a 10% possibility to generate a random room number on the digitalized floor map whenever the room detection is performed.

It is shown in Table 2 that with the additional noise the average navigation times are increased. There is little back and forth motion even when wrong detection results are presented early. But the system prompt the user to face to the doors and perform door detection more frequently in order to enhance the confidence on localization, especially when the initial localization drift is heavy. One of the issues presented is that the feature observation such as room number detection is quite slow. The room number detection is performed only if commanded by the user; and the user has to adjust his pose and stay steady before commanding.

Conclusion

In this paper, we present a wearable navigation system with context based decision making for visually impaired. By integrating multiple sensors including RGB-D camera, IMU, and web camera, the localization and trajectory of the user are functionally achieved using particle filter. We also have presented a unique approach to correct the estimated user orientation by aligning RGB-D wall border with floor map wall border, which significantly improves the fusion performance in indoor localization. The system is able to deliver semantic information to the user and help him to reach a destination. A context based decision helps to gain correct motion decision even if the temporary localization is scattered. It greatly enhances the user experience and especially fits the need of visually impaired. The future works are of two folds: one is to replace the hardware by mobile devices, and the other is to utilize more stable feature such as sparse features which minimize the need for major feature such as room number.

Acknowledgement

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 Dr. 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. Available from: <http://www.who.int/mediacentre/factsheets/fs282/en/>
2. Shim I, Yoon J. A robotic cane based on interactive technology. In IECON 02 [Industrial Electronics Society, IEEE 2002 28th Annual Conference of the]. 2002;3:2249–2254.
3. Yuan D, Manduchi R. A tool for range sensing and environment discovery for the blind. In Conference on Computer Vision and Pattern Recognition Workshop. 2004:39–39.
4. Calder DJ. Assistive technologies and the visually impaired: A digital

- ecosystem perspective. In Proceedings of the 3rdInternational Conference on Pervasive Technologies Relatedto Assistive Environments. 2010:1-8.
5. Velazquez R, Pissaloux EE, Guinot JC, Maingreud F. Walking using touch: Design and preliminary prototype of a non-invasive ETA for the visually impaired. *Conf Proc IEEE Eng Med Biol Soc.* 2005;7:6821-6824.
 6. Dakopoulos D, Bourbakis N. Wearable obstacle avoidance electronic travel aids for blind: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews).* 2010;40(1):25–35.
 7. World health organization (WHO): World report on disability. 2011. Available from: http://www.who.int/disabilities/world_report/2011/report/en/
 8. Dryanovski I, Valenti RG, Xiao J. Fast Visual Odometry and Mapping from RGB-D Data. *International Conference on Robotics and Automation (ICRA2013).* 2013:2305-2310.
 9. Foxlin E. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Comput Graph Appl.* 2005;25(6):38-46.
 10. Jimenez R, Seco F, Prieto C, Guevara J. A comparison of Pedestrian Dead-Reckoning algorithms using a low-cost MEMS IMU. In *Proc IEEE Int Symp Intell. Signal Process.* 2009:37–42.
 11. Feliz R, Zalama E, García-Bermejo JG. Pedestrian tracking using inertial sensors. *J Phys Agents.* 2009;3(1):35–42.
 12. Cheng J, Yang L, Li Y, Zhang W. Seamless outdoor/indoor navigation with WIFI/GPS aided low cost Inertial Navigation System. *Physical Communication.* 2014;13:31-24.
 13. Xiao J, Ramdath K, Losilevish M, Singh D, Tsakas A. A low cost outdoor assistive navigation system for blind people. *Industrial Electronics and Applications (ICIEA).* 2013:828-833.
 14. Mattheiss E, Krajnc E. 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.* 2013:284-295.
 15. Jeff W, Walker BN, Lindsay J, Cambias, Dellaert F. Swan: System for wearable audio navigation. In *Wearable Computers.* 2007:91-98.
 16. Yang Z, Wu C, Liu Y. Locating in fingerprint space: wireless indoor localization with little human intervention. In *Proceedings of the 18th annual international conference on Mobile computing and networking.* 2012:269–280.
 17. Kannan B, Kothari N, Gnegy C, Gedaway H, Dias MF, Dias MB. Localization, Route Planning, and Smartphone Interface for Indoor Navigation. In *Cooperative Robots and Sensor Networks.* 2014:39-59
 18. Apostolopoulos I, Fallah N, Folmer E, Bekris KE. Integrated online localization and navigation for people with visual impairments using smart phones. *ACM Transactions on Interactive Intelligent Systems.* 2014;3(4):1–28.
 19. Lan KC, Shih WY. Using smart-phones and floor plans for indoor location tracking. *IEEE Transactions on Human-Machine Systems.* 2014:1–11.
 20. Sattler T, Leibe B, Kobbelt L. Fast image-based localization using direct 2D-to-3D matching. Presented at the *Computer Vision (ICCV).* 2011:667–674.
 21. Proulx MJ, Stoerig P, Ludowig E, Knoll I. Seeing ‘Where’ through the Ears: Effects of Learning-by-Doing and Long-Term Sensory Deprivation on Localization Based on Image-to-Sound Substitution. *PLoS One.* 2008;3(3):e1840. doi: 10.1371/journal.pone.0001840.
 22. Fraundorfer F, Christopher E, Nistér D. Topological mapping, localization and navigation using image collections. *IROS.* 2007:3872-3877.
 23. Gordon NJ, Salmond DJ, Smith AF. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEEE Proceedings F - Radar and Signal Processing.* 1993;140(2):107-113.
 24. Hong S, Myeong H. Method used by robot for simultaneous localization and map-building. *US Patent.* 2011.
 25. Hu JS, Chan CY, Wang CK, Lee MT, Kuo CY. Simultaneous localization of a mobile robot and multiple sound sources using a microphone array. *Advanced Robotics.* 2011;25(1-2):135-152.
 26. Marchetti L, Grisetti G, Iocchi L. A comparative analysis of particle filter based localization methods. In *RoboCup 2006: Robot Soccer World Cup X.* 2007:442-449.
 27. Lee YH, Medioni G. A rgb-d camera based navigation for the visually impaired. *RSS 2011 RGBD: Advanced Reasoning with Depth Camera Workshop.* 2011.
 28. Joseph SL, Zhang X, Ivan D, Xiao J, Yi C, Tian Y. Semantic Indoor Navigation with a Blind-User Oriented Augmented Reality. *Systems, Man, and Cybernetics (SMC).* 2013:585-3591.
 29. Joseph SL, Yi C, Xiao J, Tian Y, Yan F. Visual semantic parameterization - To enhance blind user perception for indoor navigation. *Multimedia and Expo Workshops (ICMEW).* 2013:1-6.
 30. Yi C, Tian Y. Text Extraction from Scene Images by Character Appearance and Structure Modeling. In *Computer Vision and Image Understanding.* 2013;117(2):182-194.
 31. Tian Y, Yang X, Yi C, Arditi A. Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. *Machine vision and applications.* 2013;24(3):521-535.
 32. Kang SK, Choung YC, Park JA. Image corner detection using Hough transform. In *Pattern Recognition and Image Analysis.* 2005:279-286.
 33. Harris C, Stephens M. A combined corner and edge detector. In *Alvey vision conference.* 1988;15:50.
 34. <http://www.speech.cs.cmu.edu/pocketsphinx/>
 35. http://wiki.ros.org/audio_common
 36. Ye C, Hong S, Qian X. A Co-Robotic Cane for blind navigation. In *2014 IEEE International Conference on Systems.* 2014:1082-1087.
 37. Zhang X, Xiao J. A SLAM based Semantic Indoor Navigation System for Visually Impaired Users. *IEEE International Conference on Systems.* 2015.
 38. Lee YH, Medioni G. Wearable RGBD indoor navigation system for the blind. *Computer Vision - ECCV 2014 Workshops.* 2015:493-508.