Vision-based Mobile Indoor Assistive Navigation Aid for Blind People

Bing Li, Member, IEEE, J. Pablo Muñoz, Member, IEEE, Xuejian Rong, Qingtian Chen, Jizhong Xiao, Senior Member, IEEE, Yingli Tian, Fellow, IEEE, Aries Arditi, Mohammed Yousuf

Abstract—This paper presents a new holistic vision-based mobile assistive navigation system to help blind and visually impaired people with indoor independent travel. The system detects dynamic obstacles and adjusts path planning in real-time to improve navigation safety. First, we develop an indoor map editor to parse geometric information from architectural models and generate a semantic map consisting of a global 2D traversable grid map layer and context-aware layers. By leveraging the visual positioning service (VPS) within the Google Tango device, we design a map alignment algorithm to bridge the visual area description file (ADF) and semantic map to achieve semantic localization. Using the on-board RGB-D camera, we develop an efficient obstacle detection and avoidance approach based on a time-stamped map Kalman filter (TSM-KF) algorithm. A multi-modal human-machine interface (HMI) is designed with speech-audio interaction and robust haptic interaction through an electronic SmartCane. Finally, field experiments by blindfolded and blind subjects demonstrate that the proposed system provides an effective tool to help blind individuals with indoor navigation and wayfinding.

Index Terms—Indoor assistive navigation, semantic maps, obstacle avoidance, Google Tango device, blind and visually impaired people

1 INTRODUCTION

According to multiple federal and state civil rights laws in the United States, public areas such as airports and subway stations, need to accommodate the services and facilities accessibility for individuals with disabilities. Independent travel is always a daily challenge to those who are blind or visually impaired. According to the World Health Organization fact sheet as of October 2017, there were 6 million people who are blind and 217 million who have low vision worldwide [1].

Intelligent assistive navigation is an emerging research focus for the robotics community to improve the mobility of blind and visually impaired people. For indoor navigation on mobile devices, numerous studies have been carried out in the past decades, such as using wireless sensor network fingerprints [2], [3], [4], [5], [6], [7], [8], geomagnetic fingerprints [9], inertial measurement unit [10], and Google Glass device camera [11], [12].

There are multiple challenges for mobile indoor assistive navigation: the inaccessibility of indoor positioning, the immature spatial-temporal modeling approaches for indoor maps, the lack of low-cost and efficient obstacle avoidance and path planning solutions, and the complexity of a holistic system on a compact and portable mobile device for blind users.

The advancements in computer vision software (such as visual odometry) and hardware (such as graphics processing units) in recent years have provided the potential capabilities for vision-based real-time indoor simultaneous localization and mapping (SLAM).

As more and more indoor venue map data becomes available on public maps (e.g. Google Maps, HERE Maps and AutoNavi Maps), the integration of context-aware information becomes essential for navigation systems. User-centered ambient contextual data allows a system to anticipate a user’s personalized needs, and to customize his/her navigation experience. In this research, the indoor map with spatial context-aware information is referred to as semantic map for assistive blind navigation purposes.

In response to safety concerns, vision cameras have become more popular than other sensors (e.g. Sonar, Radar, and LIDAR) for obstacle detection and avoidance, thanks to its low cost and...
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2018.2842751, IEEE Transactions on Mobile Computing

IEEE TRANSACTIONS ON MOBILE COMPUTING

rich information in environments with proper lighting conditions. Monocular, stereo, and depth cameras have been widely used for object detection and tracking.

This paper summarizes our research work on novel intelligent situation awareness and navigation aid (ISANA) with an electronic SmartCane prototype that provides a holistic solution for indoor assistive navigation. The system uses the Google Tango tablet as its mobile computing platform.

The main contributions of this research are as follows:

1) We proposed a real-time holistic mobile solution called ISANA for blind navigation and wayfinding, which was successfully demonstrated in U.S. DOT headquarter buildings (as shown in Fig. 1), and achieved greater success than previous efforts.
2) We developed the CCNY indoor map editor to parse geometric information from architectural CAD models and extract the semantic map with a global 2D traversable grid map layer and context-aware layers which enabled the global path planning to desired destinations.
3) We presented an efficient obstacle detection and avoidance method that not only augments navigation safety but also adjusts the path in real-time.
4) We designed the CCNY SmartCane for robust HMI which mitigates the problems caused by non-perfect voice recognition software, such as malfunctions in noisy environments or command misunderstanding from users with accents.

The remainder of this paper is organized as follows. Section 2 reviews various vision-based indoor assistive navigation systems, context-aware approaches, and obstacle avoidance research. Section 3 presents an overview of the ISANA system and the high-level relationships among different modules. Section 4 elaborates the pipeline to construct semantic maps using the indoor map editor. Section 5 presents obstacle detection and motion estimation approaches using TSM-KF. Section 6 elaborates the multi-modal HMI design. Section 7 shows the quantitative and qualitative experimental evaluations. Finally, Section 8 concludes the paper and discusses the future research directions.

2 RELATED WORK

2.1 Vision-based Indoor Assistive Navigation

Advancements in computer vision provide the capabilities to support assistive navigation in real-time. In some research, vision cameras have been utilized for assistive navigation systems in several variants, including monocular [11], stereo [12], fisheye [14], omni-directional [15] and RGB-D [16].

The University of Southern California stereo system [13] was an early wearable prototype with navigation and obstacle detection functionalities. It installed a head-mounted vision sensor and a haptic feedback actuator. Stereo visual odometry was performed by matching features across two stereo views, and visual SLAM was performed subsequently. Obstacles were detected using stereo point cloud processing and were represented in the visual odometry coordinate frame. The system alerted the user when obstacles were detected in front. However, the system lacked obstacle modeling in a global map representation during navigation, such that dynamic path planning was not supported for real-time obstacle avoidance.

With rapid growth and evolution of mobile technology, mobile indoor assistive navigation systems have drawn enormous attention among researchers in recent years. The Navatar project [11] employed an inertial measurement unit (IMU) in a smart mobile device (Google Glass) for ego-motion estimation. The user of Navatar was required to confirm the presence of landmarks such as corners and doors for indoor localization. However, this user-assensor approach was undesirable since it increased the cognitive load of the blind user. Also, Navatar was incapable of performing obstacle avoidance.

The Google Tango device has been recently explored to augment indoor mobility and safety for blind and visually impaired users. Researchers at Illinois State University presented their system called Cydalion [17]. Cydalion uses depth perception to detect obstacles and head-level objects, and it provides audio tones and haptic feedback for the user. Although Cydalion makes a commendable attempt for the safety concerns using obstacle detection, it is not able to provide global navigation guidance for the user. Also, Cydalion is purely relying on the audio feedback and does not support haptic interaction. Researchers and developers from Cologne Intelligence also developed an augmented reality solution for the indoor navigation [18]. The system shows a navigation path overlaid on the 2D camera view. Nonetheless, the system is unable to perform obstacle avoidance with dynamic path planning, and there is no decent HMI for the user.

2.2 Context-aware approaches

The notion of “Context-Aware Computing” was introduced in the early 90s [19], and it was first described as an approach to customize user experience according to his/her spatial location, the collection of nearby people, hosts, accessible devices, and the changes of all these contextual information over time. The spatial reasoning was later added in the assistive navigation in an intelligent wheelchair project [20], and it used the semantic information to augment navigation and situation awareness.

A context-aware indoor navigation system was presented for user-centric path planning to help people find their destinations according to their preferences and contextual information [21], [22]. This user model included various information about physical capabilities, access rights, and user preference modeling. The system used a hybrid model of the indoor map representation which combines symbolic graph-based models and geometric models.

Knowledge representation and management for context-aware indoor mobile navigation has been presented by Afyouni [23]. The spatial query was designed and implemented using PostgreSQL database management system. Considering user-centric contextual dimensions and mobile data management issues, a hierarchical feature-based indoor spatial data model was presented by taking into account additional contextual data such as time, user profiles, user preferences and real-time events. However, this framework was not specifically designed for blind navigation.

2.3 Obstacle Avoidance Approaches

Navigation capability and safe travel are essential for indoor assistive navigation systems. Real-time obstacle avoidance is a particular challenge for autonomous robotic systems. Researchers have
explored various obstacle avoidance strategies for autonomous navigation.

As the pioneering effort of indoor robotic navigation, NavChair used sonar sensor to detect objects. A vector field histogram (VFH) algorithm \[25\] and later Minimal VFH \[26\] were proposed to provide effective sonar-based obstacle detection and avoidance. VFH used sonar readings to update a local map with detected obstacle in the form of uncertainty grids, which was used to calculate a polar histogram to represent the geometric information of the obstacles. Finally, NavChair found a direction with minimum VFH cost and followed it to bypass obstacles.

Depth maps from stereo-vision systems have been used for the detection of obstacle-free space in autonomous robotic systems. CASBliP developed a blind navigation system to fuse different information sources to help visually impaired users to deal with mobility challenges \[27\]. Based on the assumption that obstacle-free pathway is at the bottom of the depth image, CASBliP detects the obstacle-free pathway as the bottom partition of the image which has small disparities. However, this assumption does not hold for robust safety, and CASBliP is not able to perform dynamic path planning to avoid obstacles.

The RGB-D indoor navigation system from the University of Southern California for blind people can perform global obstacle avoidance during assistive navigation \[28\]. Based on the visual odometry using a glass-mounted RGB-D camera, the system first built a 3D voxel map of the environment, and a 2D probabilistic occupancy map was updated in real-time. Three grid states (occupied, free, and unknown) were represented in the global 2D traversable grid map to support dynamic path planning. Nonetheless, this traversable map lost the vertical geometric information of the obstacles. Therefore, the system was not able to detect some obstacles, such as head-level objects. Using Tango device, obstacle detection and avoidance approaches have been developed based on depth image \[29\], \[30\]. The edge detector was first used to remove all edges, and a connected component algorithm extracted all connected regions such as the floor at the bottom of the image. The non-floor image region was segmented into three sections for local obstacle avoidance guidance: bear right, bear left, and no clear path \[29\]. However, this approach can only detect and avoid obstacles locally and cannot do global path planning.

### 3 System Overview

The proposed ISANA system runs on Google Tango mobile device, which has an embedded RGB-D camera providing depth information, a wide-angle camera for visual motion tracking, and a 9-axis inertial measurement unit (IMU) for visual-inertial odometry. The physical configuration of the prototype is shown in Fig. 2 which includes a Tango tablet, a frame holder, and a SmartCane which outfits a keypad and two vibration motors on a standard white cane. ISANA provides indoor wayfinding guidance for blind users with location context awareness, obstacle detection and avoidance, and robust multi-modal HMI.

We leverage the Tango VPS and ADF for ISANA assistive navigation, as shown in the bottom right of the system functional block diagram in Fig. 3. Fast retina keypoint visual features \[31\] are extracted from the wide-angle camera for six degrees of freedom (DOF) visual odometry (VO), which is fused with IMU for visual-inertial odometry (VIO). The visual feature model is stored in ADF for loop closure detection so that the accumulated odometry drift can be suppressed.

The indoor semantic map and localization play fundamental roles for ISANA, as shown in the bottom part of Fig. 3. Based on the CAD model (or blueprint maps) as the input, the indoor map editor constructs the semantic map with multiple geometric layers (Section 4), such as global 2D traversable grid map layer and point/area of interest layers (POI and AOI), and topological layers. Then, the geometric layers are aligned with Tango ADF using the approach elaborated in Section 4.3. Therefore, ISANA achieves semantic localization to localize the user on the semantic map for path planning and location awareness.

To provide cognitive guidance for blind users, a novel way-point path planner \[32\] is implemented in ISANA based on A* algorithm and post-processing path pruning. To address safety concerns during navigation, a real-time obstacle detection approach is efficiently designed to provide the user both dynamic path planning and local obstacle avoidance using the on-board
RGB-D sensor data. Signage and text recognition is also built into ISANA for situation awareness [33].

A multi-modal human-machine interface (HMI) handles the interaction between ISANA and blind users. To cope with the conflicts of information delivery and user cognitive load, we minimize the audio announcements using event triggering. A multi-priority mechanism is designed for audio announcements to ensure higher priority messages can be conveyed on time. An electronic SmartCane is designed [34] to provide a complementary HMI through tactile input and vibration feedback, and it is especially useful in noisy environments, where speech recognition is not robust.

4 INDOOR SEMANTIC MAPS AND LOCALIZATION

4.1 Architectural Model Data

Computer-Aided Design (CAD), such as in drawing interchange format (DXF), is a common model for architectural drawings. Instead of creating all semantic information manually, retrieving geometric layers from CAD model is more efficient.

There are two fundamentally different representations for architectural data modeling: Computer-Aided Design (CAD) and Building Information Modeling (BIM). As two-dimensional electronic drawings, CAD models consist of 2D graphic elements such as lines, ellipses, and texts. Indoor BIM is a more comprehensive model that provides useful semantic information (e.g. walls, windows, rooms and space connections) that can be used for indoor navigation purposes. Although we recognize the benefit of 3D BIM data, in this paper we focus on the CAD model data since it is available for most existing buildings.

According to the CAD Standards, the DXF file includes (but is not limited to) the following entities and attributes:

1) Layering (organized in the hierarchy) and colors
2) Units and scale factors
3) Line (or polyline) types and line weights
4) Arc types and directions
5) Text style, font, and size
6) Room and door numbering
7) Title block and sheet titles
8) Drawing sequence and sheet numbers

4.2 Semantic Map Construction

We automated the map generation process by developing an indoor map editor to parse the CAD model of architectural floor map, which was provided by the CCNY Campus Planning and Facilities Management department. The map editor parses the DXF file using Dxfib library, and extracts the geometric entities of the floor’s model, including line entities, text entities, polygon entities, and ellipse entities, as shown in Fig. 5.

Fig. 5. Semantic layers extracted from CAD model using our indoor map editor. The layout is parsed from line entities, the doors are from ellipse entities, and the rooms are from text entities. Polygon entities are not applied due to it deficient representation for the hallway areas.

Further, our indoor map editor recognizes the regions of hallways, the topologies between room labels and corresponding doors, the global 2D traversable grid map layer, room area and labels, door locations. In addition, the map editor is able to edit (add or delete) any semantic information, such as connectors (e.g. elevator banks and stairs) between floors to support multi-floor
transitions. All of this spatial context-aware information is referred to as the semantic map to support assistive navigation and location awareness purposes.

To retrieve the semantic information, a lattice graph is introduced to model the architectural layout image with nodes and edges corresponding to vertices between adjacent nodes, as an undirected graph $G = (V, E)$. Then, the region growing process to find each room is modeled as a minimum spanning tree (MST) problem. MST is the spanning tree (a subgraph containing every vertex of $G$) minimizing the total weight.

As shown in Algorithm 1, ISANA first constructs the adjacency list from the global 2D traversable grid map. Next, an automated region growing approach using Prim’s algorithm is performed to find all rooms, as well as all topological connections between room labels and their connecting doors, with the computational complexity of $O((V + E) \log V)$ using adjacency list representation.

Then, the hallways are retrieved as regions which have more than certain number connecting doors (the threshold is chosen as 10 for our experiment). Finally, any other spaces which are not connecting with doors or room labels are recognized as unknown areas. The result is shown in Fig. 6.

Algorithm 1 Semantic extraction based on Prim’s MST algorithm

```plaintext
Input:
I: layout image created from CAD layers (Fig. 5)
rl: room points list
dL: door points list

Output: (Fig 6)
h: global 2D traversable grid map
g: semantic topology graph, context-aware layers

1: procedure SEMANTIC_CONSTRUCTION(I, rl, dL)
   ▷ create G as adjacency list from I
2: $G_{v,E} \leftarrow \text{adjList}(I)$,
g ← empty
3: $h \leftarrow \text{image}(I.\text{width}, I.\text{height})$ as WHITE
4: for each $s$ in rl, then in dL do
5:   $V_a \leftarrow G_{BLACK}$,
6:   $V_i[s] \leftarrow \text{WHITE}$ ▷ vertex visited array
7:   $P_q \leftarrow \text{INFINITY}$ ▷ priority queue initial value
8:   $Q \leftarrow s$ ▷ priority queue
9:   $A_d \leftarrow \text{NULL}$ ▷ adjacency list
10: while ($Q! = \text{empty}$) do ▷ Prim’s MST algorithm
11:   $m \leftarrow \text{minimum_weight}(P_q)$
12:   for $v$ in neighbor of ($m$) do
13:     update $Q, V_a, P_q, A_d$
14:     $g \leftarrow \text{edge(room, door)}$
15:     $h[s,i,s,j] \leftarrow \text{GREEN}$ ▷ room areas
16:     if $\text{door not } (g, \text{vertex}) > \text{DOOR_MIN}$ then
17:       $h[s,i,s,j] \leftarrow \text{DARK_BLUE}$ ▷ hallway areas
18:     else if $g, \text{vertex not in rl then}$
19:       $h[s,i,s,j] \leftarrow \text{RED}$ ▷ unknown areas
20: return ($g,h$)
```

4.3 Map Alignment for Semantic Localization

The map alignment and semantic localization are performed on the Tango mobile device by a sighted person. The Tango VPS provides 6-DOF pose estimation based on visual odometry and ADF features, which are understandable only by computer algorithms. However, blind users and ISANA need the semantic map to understand the scene for navigation purposes through graphical user interface (GUI) and human-machine interaction (HMI). Furthermore, the global 2D traversable grid map layer is required for the path planning algorithm.

Therefore, to establish the spatial relationship between the ADF feature map and the semantic map, we proposed a novel alignment algorithm so that ISANA is able to localize the user in the semantic map.

Fig. 7 illustrates the concept of ADF and semantic navigation map alignment method, which selects a set of control points on the semantic navigation map and the Tango device collects the ADF keyframe features in the real environment at the same physical positions as the control point set.
The alignment algorithm uses singular value decomposition (SVD) approach to find a homogeneous transformation matrix \( A \), which projects Cartesian coordinate system of the ADF to the semantic map by aligning these two corresponding point sets for optimized least-squares sum.

A SVD approach is introduced to align the ADF and the semantic map. Let us denote \( l_i = (x_i, y_i)^T \in L_{N \times 2} \) as the planar \( \mathbb{R}^2 \) coordinates of control point set from the Tango ADF local \( \mathbb{R}^3 \) coordinate system, and \( p_n = (l_n, \bar{z}_n)^T \in P_{N \times 2} \) as the corresponding point set on the semantic map. \( N \) is the number of control points, and \( n \) is the index.

The goal is to find the homogeneous transformation matrix \( A \) and it is composed by a rotation matrix \( R_{2 \times 2} \) and a translation vector \( t_{2 \times 1} \) as shown in equation (1).

\[
(R, t) \leftarrow \arg \min_{R, t} \sum_{i=1}^{N} ||R \cdot p_i + t - l_i||^2
\]

and we denote \( A = \begin{bmatrix} R_{2 \times 2} & t_{2 \times 1} \\ 0 & 1 \end{bmatrix} \) (1)

Where the \( R \in \Omega = \{R|R^T \cdot R = R \cdot R^T = I_2 \} \), and \( \Omega \) is the set of orthogonal rotation matrices.

By selecting a set of control points arbitrary distributed (rather than with a pattern like: on a line) in the semantic map, based on its corresponding control point set in ADF feature map, we have \( A \) as an over-constrained matrix, so that it guarantees the least-squares solution of equation (1).

Equation (1) is linear with respect to \( t \), but nonlinear with respect to \( R \). We denote the centroid of \( L \) and \( P \) as \( \bar{l} = \frac{1}{N} \sum_{i=1}^{N} l_i \), and \( \bar{p} = \frac{1}{N} \sum_{i=1}^{N} p_i \), and introduce the matrices \( \bar{L} = L - \bar{l} \) and \( \bar{P} = P - \bar{p} \). Therefore, the problem can be represented as the below equivalent formulation:

\[
R \leftarrow \arg \min_{R} \sum_{i=1}^{N} ||R \cdot \bar{p}_i - \bar{l}_i||^2
\]

By defining \( A \) as equation (3), we can transform equation (1) into equation (4).

\[
A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix}
\]

\[
Mx = b
\]
\[
M = \begin{bmatrix} x_1 & -y_1 & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & -y_n & 1 & 0 \\ y_n & x_n & 0 & 1 \end{bmatrix}
\]

\[
x = [\cos(\theta), \sin(\theta), t_x, t_y]^T
\]

\[
b = [l_1, l_2, \ldots, l_n, \bar{z}_n]^T
\]

Considering the cross-covariance matrix \( M \), we use the SVD to factorize it:

\[
M_{N \times 4} = U_{N \times N} S_{N \times 4} V_{4 \times 4}^T
\]

Where \( U \) includes the eigenvectors of \( M M^T \), \( S \) is diagonal matrix with singular values \( \sigma_i \), and \( V \) is the matrix with eigenvectors of \( M^T M \). Finally we find the optimal values for equation (1) by:

\[
x = (V \text{diag}(\sigma_1^{-1}, \ldots, \sigma_4^{-1})U^T)^{-1}b
\] (6)

5 Navigation with Obstacle Avoidance

In this section, we elaborate on the obstacle avoidance design to improve assistive navigation safety. Based on the global 2D traversable grid map layer in the semantic map, a global navigation graph is constructed and a path planner based on \( A^* \) algorithm is designed to generate a waypoint route for the user [36].

5.1 Obstacle Detection

The obstacle detection is performed using the RGB-D camera, and it runs efficiently in real-time on the Tango mobile device at 5 HZ.

As shown in the left part of Fig. 8, after the voxel rasterization, a denoise filter is performed to remove the outlier voxels. Rather than utilizing our previous approach [37] of random sample consensus (RANSAC) for floor segmentation, based on the roll and pitch information of the current pose, we perform an efficient de-skewing process to align the 3D voxels with the horizontal plane. Then, we applied a two 2D projection approach for obstacle avoidance.

Fig. 8. Obstacle detection and motion estimation flowchart

In the middle block of Fig. 8, the horizontal projection creates a horizontal occupancy map for dynamic path planning, and the vertical projection creates a vertical occupancy map for obstacle alerts (visualized in Fig. 14). Since this occupancy map is with time stamps in every frame, we name it time-stamped map (TSM). The whole detection process is efficient because its computational cost is determined by the granularity of the occupancy map rather than the number of the point cloud. The granularity is chosen according to the scale of the environment.

On the right part of Fig. 8, a connected component labeling approach [38] is designed for object detection on horizontal and vertical occupancy maps. Then, we choose a nearest spatial neighbor to identify each object in sequential frames. The detection process is shown in Algorithm [2]. Finally, every set of connected labels is grouped as an object.
The algorithm starts by taking the inputs of the initial estimation for two main processes: the prediction process and the update process.

5.2 Obstacle Motion Estimation

After ISANA projects the detected obstacles in the horizontal occupancy map, an efficient Kalman filter (KF) is integrated to predict the motion of each obstacle to improve the navigation safety in the dynamic environment. For a detected object in sequential frames, its state \( X \) of the discrete-time process can be estimated by the linear stochastic differential equations:

\[
X_k = A_k X_{k-1} + w_{k-1}
\]

And the measurement equation:

\[
Z_k = C_k X_k + v_k
\]

\( A_k \) Matrix \((n \times n)\) describes how the state evolves from \( k-1 \) to \( k \) without controls or noise. \( n \) is the dimension of the states.

\( C_k \) Matrix \((m \times n)\) describes how to map the state \( X_k \) to an observation \( Z_k \), where \( m \) is the vector size of the observation.

\( w_{n \times 1} \) and \( v_{m \times 1} \) are noise vectors representing the process and measurement noise respectively. The corresponding noise covariance matrices are denoted as \( Q_k = \text{cov}(w_k, w_k) = E(w_k, w_k^T) \) and \( R_k = \text{cov}(v_k, v_k) = E(v_k, v_k^T) \). The noise is assumed to be independent and normally distributed with the mean \( \mu \) at zero, as \( p(w) \sim \mathcal{N}(0, Q) \), \( p(v) \sim \mathcal{N}(0, R) \).

The KF obstacle motion estimation based on TSM consists of two main processes: the prediction process and the update process. The algorithm starts by taking the inputs of the initial estimation of the system \( X_0 \) and its associated state covariance \( \Sigma_0 \). Based on these two values, the prediction process is first utilized to predict \( X \) and \( \Sigma \). Then, the KF gain is updated and is fused with system measurement \( Z_k \) to estimate the next state \( \hat{X}_k \) and state covariance \( \hat{\Sigma}_k \).

(1) Obstacle motion prediction

The state of an obstacle is selected as:

\[
X = \begin{bmatrix} px & py & vx & vy & sx & sy & svx & svy \end{bmatrix}^T
\]

Where \( px \) and \( py \) represent the mass center of this obstacle, and \( vx \) and \( vy \) are velocities of this obstacle in the \( x \)-direction and \( y \)-direction respectively. The \( sx \) and \( sy \) are the tracking window width and length (size), and \( svx \) and \( svy \) are the size change rates.

Therefore, the prediction can be denoted as:

\[
\begin{align*}
\hat{X}_k &= A_k X_{k-1} + w_{k-1} \\
\hat{\Sigma}_k &= A_k \Sigma_{k-1} A_k^T + Q_k
\end{align*}
\]

By assuming zero mean acceleration model, the transition can be represented as equation (11).

\[
\begin{align*}
px_k &= px_{k-1} + vx_{k-1} \Delta t \\
py_k &= py_{k-1} + vy_{k-1} \Delta t
\end{align*}
\]

Therefore, we get the state transition in equation (12). \( \Delta t \) is assigned as the value of depth perception period.

\[
A_k = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & \Delta t & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\]

The Kalman gain is calculated by:

\[
K_k = \Sigma_k C_k^T (C_k \Sigma_k C_k^T + R_k)^{-1}
\]

(2) Obstacle motion correction

The obstacle position and size are measured in each camera frame, thus the observation can be represented as:

\[
Z_k = \begin{bmatrix} px & py & sx & sy \end{bmatrix}^T
\]

\[
Z_k = C_k X_k + v_k
\]

where,

\[
C_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\]

Finally, the object state vector and covariance are updated by:

\[
\begin{align*}
\hat{X}_k &= \hat{X}_k + K_k (Z_k - C_k \hat{X}_k) \\
\hat{\Sigma}_k &= (I - K_k C_k) \hat{\Sigma}_k
\end{align*}
\]

5.3 Navigation with Obstacle Avoidance

ISANA performs obstacle detection and motion estimation to improve the navigation safety. The detected obstacles are initially represented in the horizontal and vertical occupancy maps. Further, after they are detected using the connected component labeling approach, the obstacle projection in the horizontal occupancy map is transformed to 2D global traversable grid map (visualized in Fig. 14) for real-time dynamic path planning.

During navigation, ISANA detects the front obstacles and estimates the motion model in the global 2D traversable grid map. For obstacles which block the route to the next waypoint or move towards the user, our path finder algorithm updates the detected obstacles (or its predication) in the navigation graph.
and find a new path (as shown in experimental Section 7.3) to avoid the obstacles. For obstacles which are not conflicting with the planned route, ISANA only provides obstacle alerts.

6 HUMAN MACHINE INTERFACE

6.1 Operational Mode and Usability Study

ISANA works as a vision substitution for blind individuals to perform cognitive level waypoint navigation guidance. The human-machine interface (HMI) is critical for good user experience. We conducted a user preferences and needs questionnaire in our previous research [39]. The survey results from blind community show that a human-in-the-loop methodology is preferred for the assistive electronic travel aid (ETA), and query-based interaction is preferred over a menu. The subjects prefer audio as the output media for navigating guidance. The priority list of the most environmental objects to be located in indoor navigation are: doors, stairs, furniture, persons, personal items, and event text signs. Also, according to a survey made by Golledge [40], tactile feedback (vibration motors) to provide a robust HMI in noisy environments.

A human-in-the-loop HMI is designed for ISANA. The user is always in control, and ISANA adapts to the user’s action and responses according to user’s query. To effectively deliver rich semantic and guidance information while retaining robustness, a multi-modal HMI is designed for blind users with speech-audio as the main interaction modality, and with SmartCane for robust haptic interaction in noisy environments.

The operational mode of ISANA HMI is illustrated in Fig. 9. The small ellipses represent audio announcements or alerts, while the shaded command shapes indicate user inputs. ISANA HMI includes two operational modes: Idle mode (outer ellipse) and Navi-aid mode (inner ellipse). The area between outer and inner ellipses are considered to be applicable for both modes.

Fig. 9. ISANA operational mode and HMI

After the system initialization and localization from VPS, ISANA enters the Idle mode. In this mode, the user can input navigation control commands, query location awareness, and configure ISANA settings. The event alerts include obstacle in front, system exceptions (e.g. the power of the system is below a threshold) and location alerts to announce POI/AOI, and the number of alerts is minimized to decrease the cognitive load for the user. To enter Navi-aid mode, the user just needs to specify a destination.

In the Navi-aid mode, a cognitive waypoint path (how many turns to the destination) is computed by ISANA. Then, a general description of the route is announced to the user, including intermediate destinations for multi-floor routes, the number of waypoints to the (intermediate) destination and location alert (e.g. “you are in the lobby”, or “your destination is on the eighth floor”). ISANA provides verbal guidance for each waypoint and announces the distance to next waypoint in feet or steps. Waypoints are algorithmically computed and adjusted as needed when obstacles are encountered.

6.2 Speech and Audio HMI

ISANA HMI utilizes the Android text-to-audio to convey system feedback such as waypoint guidance, obstacle alerts, and location awareness information. To decrease the cognitive load for the user, an event triggering mechanism is designed to dispatch messages according to the system state machine. Meanwhile, each message is configured with a priority, so that ISANA allows higher priority messages to be delivered on time to supersede the current announcing message.

We implemented a speech-to-text voice recognition module [33] for user input by fine-tuning the CMU PocketSphinx engine. We enhanced the recognition accuracy by creating a limited vocabulary and specified grammar dictionary.

6.3 SmartCane HMI

We designed a SmartCane with tactile input (keypad) and haptic feedback (vibration motors) to provide a robust HMI in noisy environments.

Fig. 10. CCNY SmartCane installed on a standard white cane

The SmartCane unit, as shown in Fig. 10 is mounted on a standard white cane. The mechatronics functionalities of SmartCane include:

1) A SmartCane handle with a keypad and two buttons is mounted on the standard white cane to select destination candidates and confirm, and reset functionalities.
2) A 9-axis IMU is used to track the relative orientation between white cane and the user for accurate heading direction guidance.
3) The two vibration motors on the SmartCane guide the user with correct heading direction.
4) BLE is used for data transmission between ISANA and SmartCane.

http://tinyurl.com/Android-TTS
http://cmusphinx.sourceforge.net
The state machine design of the SmartCane is shown in Fig. [11]. After the SmartCane starts, it connects with ISANA via Bluetooth Low Energy (BLE) and the vibration motors are triggered to indicate a successful connection. Then, the state machine falls into the \textit{Idle} state and waits for transition events. When the user presses the keypad or any button, SmartCane delivers these inputs to ISANA.

When ISANA dispatches a heading guidance message to the SmartCane, the user can \textit{Confirm rotation} to track the relative orientation between white cane and the user. The SmartCane conveys accurate direction guidance to the user via haptic feedback. The two vibration motors indicate the directions of turning left or right, its vibration intensity decreases as the user steers to the targeted direction. Finally, vibration stops when the correct heading is reached.

7 \textbf{Experimental Results}

We tested the ISANA and SmartCane system with both blindfolded and blind subjects for its quantitative and qualitative effectiveness in guiding users to their destinations. Also, the semantic map for localization and obstacle detection modules are evaluated in details in this section.

7.1 \textbf{Map Alignment for Semantic Localization}

To verify semantic localization accuracy through the alignment, the trajectories of ISANA in Steinman Hall 6th floor (size around $35m \times 32m$) at CCNY were recorded to evaluate the closed-loop error. The closed-loop trajectory generated using pure visual odometry is drawn in red in Fig. [12] and the ISANA semantic localization trajectory based on VPS is shown in green. The ground truth is the path at the midpoint of the hallways.

The closed-loop error of both methods are shown in Fig. [13] As one can see, the pure visual odometry accumulates drift with time and leads to larger errors, whereas ISANA semantic localization performs with smaller errors and without drift.

7.2 \textbf{Obstacle Detection and Motion Estimation}

Fig. [14] shows a visualization screenshot of detected obstacles in a hallway on the ISANA App. Different objects are shown with different bounding boxes in front of the user.

The evaluation of the obstacle detection was conducted in a hallway. Fig. [15] shows the obstacles detected from the point cloud. The top two panels indicate the horizontal and vertical projections respectively. The bottom two panels show the VFH weighted histogram of both projections for obstacle-free path directions.

After obstacles are detected, a nearest-neighbor approach [41] was applied to track the objects in subsequent frames. The tracking metric is defined over the Euclidean distance between the centers of gravity of grid ensembles (shown as gray in Fig. [16] representing distinct objects.

Fig. [16] shows the result of continuous frames detection and motion estimation using TSM-KF based on field collected data. $^eP_{t-2}$ is the fused position of the object in time $t = i - 2$ which is marked in yellow. $^eP_{t-1}$ is the fused position of the object in time $t = i - 1$ shown in dark blue. $^wP_t$ shows the motion prediction position of $P_t$ shown in red. $^wP_t$ is the measurement of the $P_t$ based on the noise model.
Current pose
Detection on global 2D traversable grid map

Fig. 14. Obstacle detection results from depth perception projection in ISANA App GUI on a mobile device. In the right side, three objects (A-person, B-chair, and C-person) are projected both horizontally and vertically from the 3D point cloud. In the left side, the objects shown in green with blue outline are the detected object results.

Fig. 15. Left-top: obstacle detection horizontal projection. The horizontal axis represents the left to right direction in degree, and the vertical axis represents the forward direction in the meter. Left-bottom: obstacle histogram VFH. The horizontal axis is the angular direction in the degrees, and the vertical axis represents the weighted cost of obstacle points about their distance to the user on the horizontal plane. Right side shows the detection in the vertical projection and its VFH.

Finally, the $eP_i$ is the fused position by updating from $mP_i$ based on the prediction of $pP_i$. The circle around each position shows its covariance. The TSM-KF provides a smooth and more accurate motion estimation for obstacle avoidance.

7.3 Navigation with Obstacle Avoidance

If the detected obstacles or the predicted obstacle positions are superimposed on path to the next waypoint, ISANA generates a new path to avoid the obstacles using the path planner based on $A^*$ algorithm. Fig. 17(a) shows the path when there is no obstacle. When ISANA detects obstacles in front, a new route is generated as shown in Fig. 17(b). As the obstacle moves, route is updated to avoid the detected obstacle, as shown in Fig. 17(c).

The obstacle avoidance is running along with the assistive navigation, as shown in this demo video.

Fig. 16. Obstacle motion prediction and fusion using TSM-KF. $eP_i$ is the estimated positions (with covariance ellipsoids) of obstacle in discrete continuous frames (frame $t_{i-2}$ as shown in yellow color, $t_{i-1}$ as dark blue, $t_i$ as red). $pP_i$ shows the predicted position with covariance ellipsoid for frame $i$. $mP_i$ shows the measured position and $eP_i$ is the updated estimation using KF.

Fig. 17. Real-time obstacle avoidance during navigation

(a) The calculated route using $A^*$ and path pruning when there is no detected obstacle.

(b) When ISANA detects any obstacles in front of the user, and it superimposes with the current route, ISANA generates a new route by updating the detection into the global 2D traversable grid map.

(c) When the obstacle moves, ISANA will update the route accordingly for dynamic path planning to avoid obstacle.

4. https://tinyurl.com/ccnyisanaobstacle
7.4 System Evaluation

7.4.1 Qualitative Experiments

We initially conducted the ISANA prototyping evaluation in the Steinman Hall at CCNY with blind and blindfolded subjects.

A field demo of ISANA was further performed in U.S. Department of Transportation (DOT) headquarters buildings in Washington, D.C. in an indoor multi-floor environment. A screenshot of the test scene is shown in Fig. 18. The system successfully guided the blind subject to a specified destination with obstacle detection. The user used voice meta keyword ISANA to trigger the voice recognition and consequently specified a destination. ISANA computed the path based on the global 2D traversable grid map and guided the user through audio feedback. The field test is shown in this demo video.

Fig. 18. ISANA App GUI, screenshot from field demo at U.S. Department of Transportation (DOT) headquarter buildings in Washington, D.C.

7.4.2 Quantitative Experiments

After the DOT test, we added the SmartCane as part of ISANA for accurate heading guidance and haptic feedback. Series of field tests were conducted with blindfolded subjects in Steinman Hall at CCNY to evaluate the effectiveness of ISANA and SmartCane. The journey route from an office room to a stair (around 32 meters) takes 58 seconds for a sighted person who is familiar with the environment under a regular speed. Four blindfolded college students were selected as subjects (S1,2,3,4) for both ISANA and ISANA with SmartCane evaluation. These subjects were first trained to get familiar with the system by navigating through a different route. A tested assistive navigation journey using ISANA and SmartCane is shown in this demo video.

The experimental results were evaluated by counting walking errors (which are defined as traveling off course for more than 3 seconds) and traveling time as shown in Tab. 1. As one can see that the average (Avg.) navigation guidance error (unit: times) was reduced greatly by utilizing SmartCane for accurate heading measurement. Therefore, the traveling time (unit: seconds) was also clearly decreased for the same travel journey.

<table>
<thead>
<tr>
<th>Error(times)</th>
<th>System</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISANA only</td>
<td></td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>ISANA with SmartCane</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>1.6</td>
</tr>
<tr>
<td>ISANA only</td>
<td></td>
<td>203</td>
<td>133</td>
<td>181</td>
<td>186</td>
<td>175.8</td>
</tr>
<tr>
<td>ISANA with SmartCane</td>
<td></td>
<td>176</td>
<td>116</td>
<td>144</td>
<td>159</td>
<td>148.3</td>
</tr>
</tbody>
</table>

ISANA runs in real-time on the Google Tango mobile platforms (tested in both Tango Yellowstone and Phab 2 Pro mobile devices). Pose updating is around 100 HZ through visual-inertial odometry. The obstacle detection runs efficiently under 5 HZ during assistive navigation. The period of obstacle avoidance for path updating is adjustable to accommodate user experience, and it is set as 3 ~ 5 seconds in our experiments. Although the whole system consumes substantial power, it lasts around one and half hours while ISANA is running at full load.

8 CONCLUSIONS AND FUTURE WORK

This paper presents a mobile computing ISANA with SmartCane prototype to assist blind individuals with independent indoor travel. ISANA functionalities include indoor semantic map construction, navigation and wayfinding, obstacle avoidance, and a multi-modal (speech, audio and haptic) user interface. With vision perception situation awareness for the user’s surrounding environment and navigation aid, our experimental tests with blindfolded and blind subjects demonstrate the effectiveness of our technology as a useful tool to help blind users with indoor travel.

We conducted comprehensive experimental evaluations on blindfolded and blind subjects in various indoor environments, including both single floor and multi-floor scenarios. Our subjects learned to use ISANA fairly easily. The feedback from subjects indicates that there is still room to improve for better user interface (UI), such as: (1) the speech recognition should be improved for robustness in noisy environments; (2) the audio feedback frequency should be adjustable and customized for different users; (3) the annotation functionality of adding POI landmarks into the semantic map should be easily accessible for blind users.

Future research directions will focus on cognitive understanding and navigation in more complex and cluttered environments, such as transportation terminals.

ACKNOWLEDGMENTS

This work was supported in part by U.S. Department of Transportation (DOT) Federal Highway Administration (FHWA) grants DTFH 61-17-C-00007, DTFH 61-12-H-00002, National Science Foundation (NSF) grants CBET-1160046, EFRI-1137172 and IIP-1343402, National Institutes of Health (NIH) grant EY023483. Dr. J. Xiao thanks Google Project Tango for providing a grant to CCNY Robotics Lab as well as free Tango devices and technical support. The authors acknowledge Barbara Campbell for her valuable feedback and suggestions on ISANA, and would like to thank Dr. Ivan Dryanovsky, Dr. Chucui Yi, Dr. Samleo L. Joseph, Dr. Xiaochen Zhang, Dr. Hao Tang, Mohammed Amin, Patrick Centeno, Luciano C. Albuquerque, Norbu Tsering for their contributions to this research.
His current research interests include robotics, computer vision, machine learning, 3D semantic SLAM, assistive navigation, and robot inspection. He is a member of IEEE.

Pablo Muñoz graduated magna cum laude from Brooklyn College with a B.A. in Philosophy and membership in Phi Beta Kappa. He then went on to earn an M.S. in Computer Science from the Grove School of Engineering at the City College of New York and a Ph.D. in Computer Science from the City University of New York. He is currently a Research Scientist at Intel Corporation. His research includes the development of frameworks for video analytics that can be deployed at large scale. Previously, he has successfully led the development of localization systems for assisting visually impaired people to navigate indoors, and designed and implemented an award winning prototype to combat the spread of Zika virus using state-of-the-art computer vision techniques and citizen science.

Xuejian Rong is a Ph.D. candidate at City University of New York advised by Yingli Tian. He received the B.E. degree from Nanjing University of Aeronautics and Astronautics with honors thesis in 2013. His research interests are in visual recognition and machine learning with a focus on deep learning based scene text extraction and understanding. He also worked in the areas of image degradation removal such as image deblurring and denoising.

Qingtian Chen received his B.E. degree in Electrical Engineering from CCNY in 2017. He founded FIRST Robotics Club at CCNY to inspire others to explore the field of robotics. During his B.E. study, he conducted robotics and Robot Operating System (ROS) research in Aachen, Germany and CCNY Robotics Lab. Currently, he is working for Naval Nuclear Laboratory as an electrical engineer.

Jizhong Xiao is a Professor and Ph.D program advisor at the Department of Electrical Engineering of CCNY. He received his Ph.D. degree from the Michigan State University in 2002. M.E. degree from Nanjing Technological University, Singapore in 1999, M.S. and B.S. degrees from the East China Institute of Technology, Nanjing, China, in 1993 and 1990, respectively. He started the Robotics Research Program at the CCNY in 2002 and is the Founding Director of CCNY Robotics Lab. His current research interests include robotics and control, cyber-physical systems, autonomous navigation and 3D simultaneous localization and mapping (SLAM), real-time and embedded computing, assistive technology, multi-agent systems and swarm robotics. He has published more than 160 research articles in peer reviewed journal and conferences.

Mohammed Youssif is a Program Manager for the Accessible Transportation Technology Research Initiative (ATTRI). He is involved in research on new technology solutions for wayfinding and navigation guidance for built and pedestrian environments. He is also involved in research related to emerging technologies including wireless communications, mapping, navigation, robotics and artificial intelligence for surface transportation. Prior to joining FHWA, he worked at General Motors and Chrysler Group. He served as the expert advisor to the transportation and technology subcommittee, the national taskforce on workforce development for people with disabilities and is the co-chair of the technology subcommittee, Transportation Research Board committee on accessible transportation and mobility. He is a member of Technology for Aging Taskforce, Autism Cares Interagency Workgroup, Intelligent Robotics & Autonomous Systems (IRAS) Interagency Workgroup, Interagency Committee of Disability Research (ICDR) and Transportation Research Board committee on automated vehicles and a former member of FCC Disability Advisory Committee. He has a patent on wireless multiplex systems and methods for controlling devices in vehicle. He holds a BS in Electronics and Communication Engineering and a MS in Computer Engineering.

Aries Arditi received his Ph.D. at NYU and postdoctoral training at Northwestern University and NYU, after which he directed and conducted vision science at Lighthouse Guild and at the IBM Watson Research Center. Dr. Arditi has served as President of the International Society of Low Vision Research and Rehabilitation, as Editor-in-Chief of the Journal of Visual Impairment and Rehabilitation, and on numerous U.S. government committees and panels relating to vision science. He is a research diplomat and fellow of the American Academy of Optometry and a fellow of the American Psychological Society. He is author of over 150 scientific publications. With wide-ranging interests in vision science, accessibility and vision rehabilitation, he now conducts research at Visibility Metrics LLC, and develops clinical vision tests through the Mars Perceptrix Corporation.

Yingli Tian (M99SM01F18) received the B.S. and M.S. degrees from Tianjin University, China, in 1987 and 1990, and the Ph.D. degree from Chinese University of Hong Kong, Hong Kong, in 1996. After holding a faculty position at National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, she joined Carnegie Mellon University in 1998, where she was a postdoctoral fellow at the Robotics Institute. She then worked as a research staff member in IBM T. J. Watson Research Center from 2001 to 2008. She is one of the inventors of the IBM Smart Surveillance Solutions. She is a professor in the Department of Electrical Engineering at the City College and the department of Computer Science at the Graduate Center, City University of New York since 2008. Her current research focuses on a wide range of computer vision problems from object recognition, scene understanding, to human behavior analysis, facial expression recognition, and assistive technology. She is a fellow of IEEE.