ARMSAINTS: An AR-based Real-time Mobile System for Assistive Indoor Navigation with Target Segmentation

Jin Chen, Arber Ruci, E'edresha Sturdivant, and Zhigang Zhu

Abstract— **This paper proposes an AR-based real-time mobile system for assistive indoor navigation with target segmentation (ARMSAINTS) for both sighted and blind or low-vision (BLV) users to safely explore and navigate in an indoor environment. The solution comprises four major components: graph construction, hybrid modeling, real-time navigation and target segmentation. The system utilizes an automatic graph construction method to generate a graph from a 2D floorplan and the Delaunay triangulation-based localization method to provide precise localization with negligible error. The 3D obstacle detection method integrates the existing capability of AR with a 2D object detector and a semantic target segmentation model to detect and track 3D bounding boxes of obstacles and people to increase BLV safety and understanding when traveling in the indoor environment. The entire system does not require the installation and maintenance of expensive infrastructure, run in real-time on a smartphone, and can easily adapt to environmental changes.**

I. Introduction

In recent years, the demand for indoor navigation services has rapidly increased with the increasing development of large-scale infrastructure projects [1]. The complexity and crowdies throughout large facilities (e.g., shopping malls, transportation centers, and hospitals) bring more challenges for people to determine their positions and paths to their destinations. People who are blind or have low vision (BLV) face more issues in understanding and traveling in such complex dynamic environments. BLV can only rely on their limited-visual sources, such as people assistance, white canes, guide dogs and/or touch objects, to become familiar with their surroundings. It is not only time-consuming for them to approach their destinations, but also has safety concerns for independent travel.

Indoor positioning and navigation applications have been widely studied and developed with the increase in public demand and the availability of micro-electromechanical systems (MEMS) sensors and deep learnings techniques. These techniques also applied to the service robots to perform various tasks [2, 3], such as food delivery and object management in warehouses. Most existing indoor position systems utilize radio frequency identification tags (RFIDs) [4, 5], wireless fidelity (WiFi) [6, 7], Bluetooth sensors [8, 9, 10, 11] and ultra-wideband (UWB) [12]. However, these methods do not support precise localization or require high costs for pre-installed infrastructures. Additionally, existing systems often do not consider the need for BLV or people with other disabilities, such as autism spectrum disorder, who might need to avoid the crowds and obstacles along the ways [13, 14].

Instead of using a mobile robot for assisting BLV or even sighted people in the navigation of an indoor environment, this paper proposes a solution of integrate the mobility of humans and the semantic and navigation capability of an intelligent mobile app to form a more portable "robotic" solution for a greater social impact with more affordable hardware. We developed an augmented reality (AR)-based real-time mobile system for assistive indoor navigation with target segmentation (ARMSAINTS), which utilizes ARKit and provides personalized turn-by-turn navigation instructions along with obstacle avoidance. The system has a simple scanning process minimizing the need for installation and maintenance of expensive infrastructure, eliminating the need of a mobile robotic platform if carried by a person, but has the potential to apply to service robots. This paper is an extension of our previous work [15], and the key contributions include: (1) A method that automatically constructs graphs from 2D floorplans. (2) The Delaunay triangulation-based method improves the localization accuracy. (3) A real-time 3D obstacle detection method integrates AR-based 3D models with 2D detection and segmentation models.

The remainder of this paper is organized as follows. Section II discusses the state-of-the-art indoor navigation techniques. Section III provides an overview of the system architecture and its main functionalities. Section IV describes the system data collection and integration process with graph construction and version control method. Section V demonstrates how the system using the collected data provides real-time assistive navigation service along with obstacle detection to enhance the semantic target segmentation from the surrounding scene. Section VI presents experimental results and discusses the potential applications of our system. Finally, we provide conclusions and future research directions in Section VII.

II. Related Work

A. WiFi-Based Indoor Position Systems

WiFi positioning techniques are commonly used in indoor localization, with WiFi infrastructure available across most

* The work is supported by NSF (#2131186, #2118006, #2048498, #1827505, #1740622, and #1737533), Intelligence Community Center for Academic Excellence (IC CAE) at Rutgers (#HHM402-19-1-0003 and #HHM402-18-1-0007) and AFOSR (#FA9550-21-1-0082).

J. Chen is with the Department of Computer Sciences, The City University of New York - CUNY, New York, NY 10031 USA, and also with Nearabl Inc., New York, NY 10023 USA (e-mail: jchen025@citymail.cuny.edu).

A. Ruci is with NY I-Corps Hub – CUNY, New York, NY USA, and also with Nearabl Inc., New York, NY 10023 USA (e-mail: arber.ruci@cuny.edu).

E. Sturdivant is with Nearabl Inc., New York, NY 10023 USA (e-mail: eedresha@nearabl.com).

Z. Zhu. is with the Department of Computer Sciences, The City University of New York - CUNY, New York, NY 10031 USA, and also with PhD Program in Computer Science, The Graduate Center - CUNY, New York, NY 10016 USA (e-mail: zhu@cs.ccny.cuny.edu).

indoor areas. Signal strength-based methods use trilateration techniques to compute user locations with different algorithms using the received signal strength indicator (RSSI) and other information (e.g., signal arrival time) from at least three wireless access points (APs) [16]. However, localization accuracy is highly dependent on the number of APs installed. Fingerprint-based methods [17] provide higher localization accuracy than signal strength-based methods; however, dynamic changes in access points and the environment affect the fingerprint values at the corresponding locations, which would require an update in the fingerprint database. To adapt to the dynamic changes in fingerprints, [6] proposed a maximum feature adaptive online sequential extreme learning machine that uses transfer learning to preserve previous knowledge of individual motions in the incremental learning model. Deep neural networks also apply to improve location estimation; such that Qin et al. [7] used a convolutional denoising autoencoder and convolution neural network to extract and train key features of RSSI values. Although many studies have improved WiFi-based localization techniques, they still have positioning errors between 2 and 4 meters.

B. Bluetooth-Based Indoor Position Systems

Bluetooth sensors have lower cost and power consumption than WiFi APs. Similar to WiFi APs, Bluetooth sensors also emit radio frequency signals that can be used to calculate the distances between the sensors and the user device location. Bluetooth sensors typically have two approaches for positioning. The first is the proximity and RSSI ranging techniques, such as the trilateration method [8] and log-distance path loss model [9], which are used to determine the approximate regions that user or object locations and require fewer installed sensors. The second approach uses geometric calculations to determine more precise user locations. Murata et al. [10] improved the probabilistic localization algorithm using a particle filter, motion model and observation model to solve the problems of varying RSS values due to inflation, signal delays, and continuous localization for different walk patterns, etc. The proposed method decreased the mean localization error from 3m to 1.5m through experiments in shopping malls of 21,000 m^2.

C. Marker-Based Indoor Position Systems

Visual positioning systems also do not require the installation of expensive infrastructure. The system estimates the device location based on the visual angle from the device to predefined markers. NaviLens [18] is a navigation application that detects unique QR codes that embed navigation information to assist BLV users in traveling. They optimized the reading capability of QR codes to allow detection from a distance of 15m. Similarly, [19] used QR codes to provide localization and navigation services to attender robots using contour detection techniques. However, light factors affect the QR code decoding process and obstacles along the way can lead to collisions.

Augmented reality markers are more accurate than QR codes and are more flexible for marker appearance. Sato [20] utilized AR markers with BLE sensors for user localization with power-saving mode. Consequently, that localization error was within 30 cm. However, this marker-based positioning system still requires the installation and maintenance the markers and also affected by the light factors.

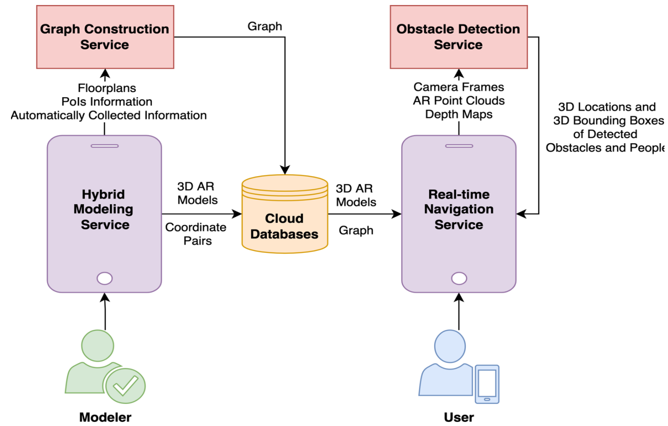


Figure 1. The ARMSAINTS system overview.

Our proposed assistive indoor navigation system minimizes the maintenance of pre-installed infrastructure, as it models the existing environment visual features with ARKit [21] and uses it to determine the initial user localization. ARKit's world tracking functionality, built on top of the visual-inertial odometry [22] technique, is used to track user locations. Our proposed region segmentation and transition methods [15] overcome the limitations of large cumulative localization error and memory constraints for long-distance tracking in AR-based position methods.

III. System Overview

The AR-based assistive indoor navigation system ARMSAINTS (Fig. 1) augments users' understanding of their surroundings and provides real-time turn-by-turn guidance with highly accurate localization and multimedia interactions. The entire system was separated into two stages: data collection and integration, and real-time assistive navigation. Data collection and integration consists of a graph construction service and a hybrid modeling service, whereas real-time assistive navigation includes a real-time navigation service and a target (obstacle and people) segmentation service.

Graph Construction Service: This service processes building floorplans (blueprints) and generates all the walkable paths with an automatic graph construction algorithm. It integrates the point of interests (PoIs) information provided by a modeler (or a facility manager) and other automatically collected information (e.g., data connectivity maps) during scanning to construct the spatial database of buildings that will be used by the real-time navigation service to provide guidance for users. The details of the automatic graph construction method are described in Section IV.

Hybrid Modeling Service: An iOS app is created for the modeler to scan spatial features of the environment with ARKit and record corresponding coordinate pairs between the scanned 3D AR model and floorplans, along with other automatically collected information. The scanned AR models and their corresponding coordinate pairs are used as bases for user localization. The hybrid modeling service utilizes our region segmentation and transition methods [15], Delaunay triangulation-based localization and version control method to achieve highly accurate user localization and easy adaptation to environmental changes.

Real-time Navigation Service: An iOS app is created for both sighted and BLV users to explore or navigate scanned facilities. It takes user requests for destination information and identifies or ranks similar destination landmarks. The real-time navigation service utilizes the spatial database and 3D AR models to determine and track real-time user locations. It uses our personalized route planning algorithm [15] to determine the best traversable navigation path based on the user preferences, real-time building status, and other factors.

Target Segmentation Service: This service is responsible for detecting the obstacles and people appearing in the view of the user's camera and determine their 3D locations and bounding boxes. It is mainly used to help BLV users avoid obstacles and augment their understanding about and interaction with their surroundings. The service also utilizes an information filtering module to determine the feedback messages based on user preferences.

IV. System Data Collection and Integration

Accurate user localization and personalized path planning are the two essential functionalities of assistive navigation services. The system utilizes a region-based modeling method [15] for localization that allows for lazy-loading and lowers the computational requirement with a small set of geolocation features during the localization matching process. This method generates a 3D AR model for each region of the building, and each AR model has a unique coordinate system that needs to be matched individually with the corresponding building floorplans. To enable the localization service, the system must contain building floorplans, 3D AR models and associated transformation matrices to map the coordinate system of the models and building floorplans. To improve the localization accuracy, the Delaunay triangulation was applied to the floorplan traversable regions to compute transformation matrices.

Various environmental variables can influence users' navigation experience in complex environments [23], such as floorplan configuration, signage information, architectural settings, and users' physical and cognitive status. Multiple environmental information, such as data connectivity maps, obstacle information and PoI information, is collected and integrated to construct the graph of the building and used by our personalized path planning algorithm [15] to obtain the most suitable paths for each user. The original graph construction method [15] was built on top of the modeler's walking path during scanning process, however, this method has many limitations. It required the modeler to travel all the walkable paths of the corresponding region during scanning, and often require modeler to travel a path segment repeatedly to cover all the walkable paths. In addition, there are problems with the landmark connections between the nearby regions. For example, connecting the first and last landmarks of two consecutive model regions can lead to duplicate landmarks in the overlapping area and not always provide an optimal path. Replacing nearby landmarks might introduce errors in creating incorrect connections between landmarks. To mitigate this problem, we developed an automatic graph construction method to extract walkable paths from 2D building floorplans and integrate them with the collected information from modeling and user input.

A. *Delaunay Triangulation-based Localization*

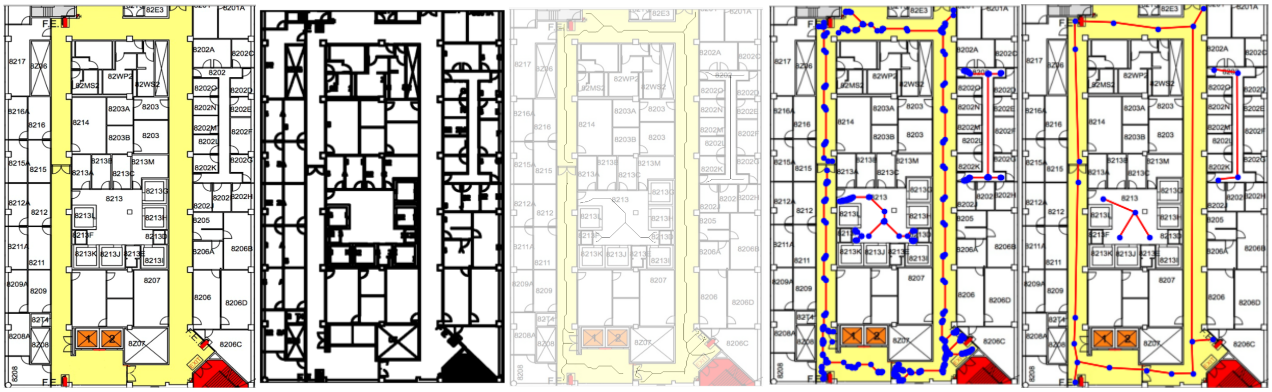
The hybrid modeling service collects multiple coordinating pairs during the scanning stage with the modeler input. To improve localization accuracy and minimize human error during the manually indicated coordinate pairs, we utilized Delaunay triangulation to divide the scanned region into multiple triangular areas based on the collected coordinating pairs. Instead of a single alignment transformation matrix generated by all the corresponding pairs, each triangular area would have its own alignment transformation matrix that maps the region model and the building floorplans' coordinate system with the three coordinating pairs of vertices using affine transformation.

The computed transformation matrices and coordinate pairs are stored in the database, along with the AR model. To determine the location of users in the floorplan from the AR camera coordinates, the corresponding coordinate pairs are used to find the associated triangular area or the nearest triangular area, the corresponding transformation matrix is used to convert the camera coordinates into the floorplan coordinates. A similar process was used to convert the floorplan coordinates back to the AR camera coordinates using an inverse transformation matrix.

B. *Automatic Graph Construction*

The system includes an automatic graph construction method that takes the 2D building floorplan and returns the graph containing a set of nodes and edges, where each edge is considered a traversable path between the associated nodes. To extract the graph from the floorplan image, the automatic graph construction method first removes the text from the image using a pre-trained deep learning model, as the text does not belong to the building infrastructure and is considered as noise. Next, the image is converted into grayscale and applied Otsu's thresholding to minimize foreground noise that does not belong to building infrastructure. Subsequently, the algorithm will cover the small objects (e.g., tables and stairs) by adaptive thresholding based on image size and floorplan scale and thicker lines in the image to decrease the noise in path extraction, as shown in Fig. 2 (b). The resulting image is converted into a binary image, and the skeletonization process that utilizes the hit-and-mass transformation is applied to extract the central path, as shown in Fig. 2 (c).

The algorithm then extracts the pixel locations of the ending points of lines and connected turning points between lines, treats them as the nodes in the graph and creates the associated edges, as shown in Fig. 2 (d). The resulting graph contains noise due to the inconsistent pixel locations of the lines, especially for the curve paths; therefore, a graph cleaning process will be applied to remove the noise and save the final graph (Fig. 2 (e)) to the database. The extracted graph from the image is used as the base and is then integrated with the data connectivity maps collected during the scanning stage and PoI information from the modeler input to construct the complete graph for our personalized path planning algorithm. When a new region is added to the floorplan or an update in the region boundary, the graph will also be processed to update the associated regions for each node in the graph.



(a) (b) (c) (d) (e)

Figure 2. Automatically graph construction process. (a) original floorplan; (b) cleaned binary floorplan; (c) extract path with the floorplan; (d) rough nodes in blue dots and edges in red lines extracted from the path; (e) result clean graph of the floorplan path.

C. Version Control Method

Indoor environments are constantly changing, such as new decorations, adding new infrastructures, or adjusting architecture design. These changes often affect WiFi-based and Bluetooth-based localization techniques, as they influence the inflation of signal strength. Our AR-based localization technique is also impacted by these changes as the spatial features are different from the recorded features in the previous scanned AR model; thus, it cannot match the user location. Moreover, such changes can lead to update in floorplan and require new walkable paths. We developed a version control method to handle these changes with our region-based modeling method.

In the case of a major change with an update in the floorplan, a new graph is generated using the automatic graph construction method and PoI information is maintained for the part of the unchanged floorplan area. The regions of the corresponding floor with changes are removed, and a new modeling process is required. However, if it only decoration changes or minor updates that do not require floorplan updates, then only the associated regions require rescanning. Additionally, as our AR-based localization method is based on feature matching, the lightning conditions or time-specific layout can introduce different spatial environments at different times. The system allows multiple AR models for same region and uses the corresponding AR model based on the condition.

V. ASSISTIVE NAVIGATION SERVICES

The system supports both exploration and navigation modes. In the exploration mode, users walk around and receive information about their surroundings, including PoI information and feedback messages from the target segmentation service, based on user preferences. In the navigation mode, users can enter their desired destination information, and the system searches and ranks similar destination locations with different constraints or users can directly select the destination from the floorplans. With a selected destination, the personalized path planning algorithm will provide the best route and the navigation app will provide turn-by-turn guidance to help users travel to their destinations. The obstacle detection service will be performed simultaneously if needed to help BLV users to avoid obstacles along the route or provide additional semantic information such as the presence of people and other targets to enhance their understanding of the surroundings.

A. Obstacle and People Detection

Target segmentation is not only important for BLV users to avoid obstacles along the path but also to help them with semantic understanding and even interact with others. Instead of creating a new system for obstacle and people detection, it is built on top of the ARKit capability. The obstacle and people detection method utilizes a 2D object detector (i.e., YOLOv3 [24]), semantic segmentation model (i.e., DeepLabv3 [25]), 2D object tracking method, 3D ARKit point cloud, and depth map generated by LiDAR sensors in the recent iPhone versions.

The method starts by detecting the 2D object bounding boxes and associated labels using a 2D object detector with a capture camera frame. ARKit contains the point cloud for each frame. It is projected onto the corresponding 2D image coordinate system to estimate the 3D bounding boxes of each detected objects from the 2D detection results using grouping.

The ARKit point cloud is sparse and requires validation across multiple frames to obtain stable 3D bounding boxes of the objects using a non-maximum suppression method. Nevertheless, it is very time efficient and require low computation power. Due to the dynamic nature of people, this method is not suitable for determining and tracking their 3D locations and bounding boxes overtime. Therefore, the 3D bounding box estimation process is different for detected people. People location and bounding box estimation applied semantic segmentation for each detected 2D people bounding box and then matched the segmented region with the depth map obtained through the LiDAR sensors embedded in the iPhone to determine the people' 3D bounding boxes. The method also uses the plane detection capability of ARKit to detect objects with planar surfaces, this covers large objects that cannot be detected by the 2D object detector.

Once the system has detected objects' bounding boxes and labels, it activates the tracking mode to track the 2D object bounding boxes over frames and uses the above methods to update the objects' 3D bounding boxes. The 2D object detector is reactive when the object is loss tracked in the frames. A 3D object matching module is applied to match the objects between the tracking mode and the detection mode to avoid duplicating the same object.

Finally, the information filtering module handles the audio/tactile feedback messages to the user for a massive obstacle detected over time based on user preferences. It consists of two modes, obstacle avoidance and semantic understanding. For obstacle avoidance, it simply provides the information about moving directions (e.g., slightly left, slightly right, and straight) to avoid obstacles along the way. For semantic understanding, it provides information about the detected targets (e.g., target labels and sizes, distance away.) based on user settings and the angular range for the targets to be considered in the feedback. AR visualization of detected targets can also be added to provide additional visual aid for people with low vision or other impairments based on user preferences.

B. *Destination Search and Navigation*

With the region-based modeling method, various techniques can be employed to determine the region that user is located, as precise accuracy is not required in this case. These techniques include but are not limited to QR codes, BLE beacons and the WiFi positioning system, based on the availability of existing building facilities. With the determined region, the corresponding AR model is loaded to determine and track the precise user locations using the Delaunay triangulation-based localization described previously and the region transition method [15].

Finding the desired destination in an indoor environment could be more challenging than outdoors, such that it does not have a common address or name. To help users better locate their destinations, we developed a search ranking algorithm that ranks the destinations based on the string-matching scores of the destination name, distance between the destination and user location, and accessibility between floor transitions based on user physical status. The navigation app receives both text and audio inputs from users, and users can also directly select the destination from the floorplans. With the selected destination, a suitable path is computed using a personalized path planning algorithm. The navigation app will then provide turn-by-turn guidance to assist users to go to their destinations along with the visual aids of AR arrows if needed. The information of nearby PoIs will also be either visualized with AR or will provide audio feedback based on user settings to help users better understand their surroundings.

VI. Results and Discussions

The proposed assistive indoor navigation system ARMSAINTS was tested with several polit projects, including a school campus, a food production site, an office building, a construction site, among others. Some key information is listed in Table I. In one example, it took approximately four hours to set up an area of 270 m^2 with given floorplans, the setup including region segmentation and modeling, entering PoI information, uploading pre-collected images or videos of the area, and setup region determination methods (e.g., place QR codes or BLE beacons). After the setup, users can use the navigation app to explore and navigate the scanned area as expected by receiving the messages about their surroundings.

TABLE I. Experiment Sites Details

Location Type	Area (m^2)	Number of Regions	Number of Landmarks
School Campus	270	2	15
Food Production Site	420	3	12
Office Building	650	2	18
Construction Site	2700	6	55

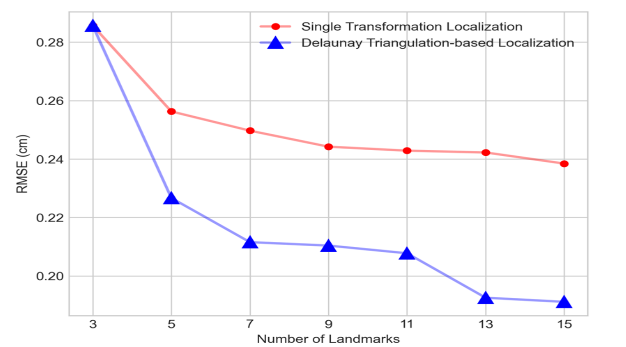


Figure 3. Comparsion of localization accuracy for 36 test points using single transformation and delaunay triangulation-based localization.

A. *Experiment Results*

To evaluate the localization accuracy, a total of 36 ground truth points were selected in a 60 m^2 test area. The test area was first scanned with recorded 15 landmarks, and then a sighted user stood at the 36 locations and used our app to estimate their positions. Fig. 3 shows the experimental results, where the Delaunay triangulation-based localization method was employed with increasing numbers of control points, showing that more accurate results with more landmarks used. We also tested at different pilot locations as listed in Table 1, where the Delaunay triangulation-based localization method minimizes cumulative errors in large area regions.

The obstacle detection method is also being tested, and can achieve a time performance of 10-15 fps with an iPhone 13 Pro Max. The variation in fps is due to the sparsity of the AR point cloud of each frame which does not always contain sufficient feature points to estimate 3D bounding boxes of the detected objects. The vacuum robot in Fig. 4 (a) is not included in the object classes in the 2D object detector (YOLOv3) but can be detected with ARKit plane detection and considered as an obstacle. The person and chair were detected using YOLOv3, and 3D bounding boxes were successfully obtained using the obstacle detection method. Because bottom part of the person is occluded by the chair, its 3D bounding box can only cover the visible part. However, it was sufficient to support the functionalities of our system, which is to help BLV users avoid obstacles and provide a semantic understanding of their surroundings.

B. *Potential Applications*

With the highly accurate localization, personalized path planning and obstacle detection method, the system can also be applied to service robots in additional to the current

(a) (b)

Figure 4. 3D Obstacle Detection Results. (a) detected vacuum robot as obstacle; (b) detected person and chair.

assistive navigation application with human-robot collaboration. All we need is a plugin slot for a smartphone to robot platform and the communication between them can be Bluetooth. It can help to accomplish dedicated tasks with robots, such as food delivery in restaurants or hotels, scheduling floor cleaning with vacuum robots, and objects arrangement in warehouses. Moreover, with people detection and real-time user localization, it is also possible to allow robots to deliver items not only to a desired destination but also to a specific person. Furthermore, with a personalized path planning algorithm, it could also allow the robot to perform floor transition if elevators are available and can further integrate with traffic control methods to better manage traffic flow.

VII. Conclusion

This paper proposes an assistive indoor navigation system, that can assist both sighted users and BLV users to explore and navigate the indoor environment with advanced semantic information about the surroundings. The region-based modeling method and Delaunay triangulation-based localization method provides precise localization with low-cost infrastructure. The version control method allows easy adaption to environmental changes. Furthermore, the obstacle detection service increases the safety of BLV users during indoor travels and provides them with additional information for them to better understand their surroundings. The system has proven its feasibility through experiment with multiple polit projects. Additionally, the system has the potential to be integrated with service robots to provide localization and navigation services that assist with delivery tasks to both static locations and dynamic people locations. In the future, we intend to improve the automatic graph construction method by detecting different elements of the floorplan (e.g., doors, stairs, and mechanical equipment) to construct a better graph. Furthermore, developing an automatic scanning process with robots can minimize localization error that cause by the human error in indicating coordinate pairs between AR model and floorplan coordinates.

References

[1] N. El-Sheimy and Y. Li, "Indoor navigation: State of the art and future trends," *Satellite Navigation*, 2(1), 1–23, 2021

[2] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman, "Analysis and observations from the first amazon picking challenge," *IEEE Trans. ASE,* 15(1), 172–188, 2016.

[3] Y. Sun, L. Guan, Z. Chang, C. Li, and Y. Gao, "Design of a low-cost indoor navigation system for food delivery robot based on multi-sensor information fusion," *Sensors*, 19 (22), p. 4980, 2019.

[4] C. Tsirmpas, A. Rompas, O. Fokou, and D. Koutsouris, "An indoor navigation system for visually impaired and elderly people based on radio frequency identification (rfid)," *Info. Sci.*, v 320, 288–305, 2015.

[5] C. Li, L. Mo, and D. Zhang, "Review on uhf rfid localization methods," *IEEE J. Radio Freq. Identification*, 3(4), 205–215, 2019.

[6] A. S. Al-Khaleefa, M. R. Ahmad, A. A. M. Isa, A. Al-Saffar, M. R. M. Esa, and R. F. Malik, "Mfa-oselm algorithm for wifi-based indoor positioning system," *Information*, 10(4), p. 146, 2019.

[7] F. Qin, T. Zuo, and X. Wang, "Ccpos: Wifi fingerprint indoor positioning system based on cdae-cnn," *Sensors*, 21(4), p. 1114, 2021.

[8] A. Noertjahyana, I. A. Wijayanto, and J. Andjarwirawan, "Development of mobile indoor positioning system application using android and bluetooth low energy with trilateration method," IEEE ICSIIT, pp. 185–189, 2017.

[9] A. Satan and Z. Toth, "Development of bluetooth based indoor positioning application," IEEE Future IoT, pp. 1–6, 2018.

[10] M. Murata, D. Ahmetovic, D. Sato, H. Takagi, K. M. Kitani, and C. Asakawa, "Smartphone-based localization for blind navigation in building-scale indoor environments," *Pervasive and Mobile Computing*, v 57, 14–32, 2019.

[11] D. Sato, U. Oh, J. Guerreiro, D. Ahmetovic, K. Naito, H. Takagi, K. M. Kitani, and C. Asakawa, "Navcog3 in the wild: Large-scale blind indoor navigation assistant with semantic features," *ACM Trans. ACCESS*, 12(3), 1–30, 2019.

[12] A. Poulose, O. S. Eyobu, M. Kim, and D. S. Han, "Localization error analysis of indoor positioning system based on uwb measurements," IEEE *11th Int. Conf. on Ubiquitous &Future Networks.* 84–88, 2019.

[13] A. D. Smith, "Spatial navigation in autism spectrum disorders: a critical review," *Frontiers in Psychology*, v 6, p. 31, 2015

[14] N. A. Giudice, "Navigating without vision: Principles of blind spatial cognition," in *Handbook of behavioral and cognitive geography*. Edward Elgar Publishing, 2018.

[15] Z. Zhu, J. Chen, L. Zhang, Y. Chang, T. Franklin, H. Tang, and A. Ruci, "iassist: An iphone-based multimedia information system for indoor assistive navigation," *Int. J. of Multimedia Data Engineering and Management (IJMDEM),* 11(4), 38–59, 2020.

[16] B. Choi, K. La, and S. Lee, "Uwb tdoa/toa measurement system with wireless time synchronization and simultaneous tag and anchor positioning," *IEEE Int. Conf. CIVEMSA*, pp. 1–6, 2018.

[17] J. So, J.-Y. Lee, C.-H. Yoon, H. Parket al., "An improved location estimation method for wifi fingerprint-based indoor localization," *Int. J. Software Engineering and Its Applications*, 7(3), 77–86, 2013.

[18] Neosistec, *The cutting edge technology for the visually impaired*, NaviLens. Accessed on: Feb. 22, 2022. [Online] Available: https://www.navilens.com/en/#navilens-section

[19] A. Sneha, V. Teja, T. K. Mishra, and K. N. Satya Chitra, "Qr code based indoor navigation system for attender robot," *EAI Endorsed Transactions on Internet of Things*, 6(21),2020.

[20] F. Sato, "Indoor navigation system based on augmented reality markers," *Int. Conf. Innovative Mobile and Internet Services in Ubiquitous Computing*. Springer, pp. 266–274, 2017.

[21] Apple Inc., *Augmented reality*, Apple Inc. Accessed on: Feb. 22, 2022. [Online] Available: https://developer.apple.com/augmented-reality/

[22] V. Usenko, J. Engel, J. St¨uckler, and D. Cremers, "Direct visual-inertial odometry with stereo cameras," *ICRA*, 1885–1892, 2016.

[23] J. Weisman, "Evaluating architectural legibility: Way-finding in the built environment," *Environment and behavior*, 13(2), 189–204, 1981.

[24] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[25] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.