Article

AR-Based Navigation Using RGB-D Camera and Hybrid Map

Woranipit Chidsin, Yanlei Gu * and Igor Goncharenko

College of Information Science and Engineering, Ritsumeikan University, Shiga 525-8577, Japan;
woranipit.c@gmail.com (W.C.); igor@fc.ritsumei.ac.jp (I.G.)
*   Correspondence: guyanlei@fc.ritsumei.ac.jp

Abstract: Current pedestrian navigation applications have been developed for the smartphone platform and guide users on a 2D top view map. The Augmented Reality (AR)-based navigation from the first-person view could provide a new experience for pedestrians compared to the current navigation. This research proposes a marker free system for AR-based indoor navigation. The proposed system adopts the RGB-D camera to observe the surrounding environment and builds a point cloud map using Simultaneous Localization and Mapping (SLAM) technology. After that, a hybrid map is developed by integrating the point cloud map and a floor map. Finally, positioning and navigation are performed on the proposed hybrid map. In order to visualize the augmented navigation information on the real scene seamlessly, this research proposes an orientation error correction method to improve the correctness of navigation. The experimental results indicated that the proposed system could provide first-person view navigation with satisfactory performance. In addition, compared to the baseline without any error correction, the navigation system with the orientation error correction method achieved significantly better performance. The proposed system is developed for the smart glasses and can be used as a touring tool.

Keywords: navigation; augmented reality; SLAM; map

1. Introduction

In recent years, pedestrian navigation has become one of the most important services in people's city lives. Now, people are increasingly turning to their mobile phones for maps and directions when on the go [1]. As the smartphone is still the most widely used portable device, the current pedestrian navigation applications have been developed for the smartphone platform and guide users from the 2D top view map. However, with the development of smart glasses, more and more scientists and engineers started to move their attention to Augmented Reality (AR)-based navigation from the traditional top view-based navigation [2–4]. In the AR-based navigation system, the navigation information is integrated with the real scene and visualized from the first-person view. The convenience and reality can be drastically improved in the AR-based navigation compared to the current navigation apps and using paper maps [5]. In addition, most of the pedestrian navigation applications are developed for outdoor navigation, there are few systems that can be used for an indoor environment. This paper proposes an AR-based navigation system for an indoor environment.

The quality of the navigation service highly depends on the accuracy of the positioning. Global Navigation Satellite System (GNSS) is the most developed positioning system in the world and it is widely applied in outdoor navigation. However, GNSS does not work very well inside a building because the structures interfere with the signal [6]. Considering the available sensors in the current smart devices, Wi-Fi, Bluetooth, Pedestrian Dead Reckoning (PDR), and camera are possible solutions for the indoor positioning and navigation.

Wi-Fi-based method is a popular choice for indoor positioning. With communication technology advancing with every passing moment, the Wi-Fi access points, whether it be a public or private network, are increasing rapidly. On top of having more access

points (APs), portable Wi-Fi chips are found on most daily driver consumer electronics, for example smartphones, portable music players, smartwatches, smart glasses, or other smart devices [7]. There are two main groups of Wi-Fi positioning methods: fingerprint-based method and trilateration-based method. Fingerprint-based method needs to create a radio map of many calibration points in the offline phase. Positioning is conducted by estimating the similarity between the received Wi-Fi fingerprint and the fingerprint pre-stored in the radio map [8,9]. The more accurate positions of the calibration points in the radio map allow the benefit of better positioning results [10]. The trilateration-based method works well in a line-of-sight environment [11], and the accurate location information of each AP is needed [12]. Both approaches have similar problems. It is very time-consuming to create a high-quality Wi-Fi access point database and fingerprint database.

Bluetooth is also another possible solution; similar to the Wi-Fi approach, there are many devices capable of receiving Bluetooth. Compared to the Wi-Fi method, Bluetooth beacons used less power [13], but they are not so widely deployed as Wi-Fi routers. Moreover, the access point database or fingerprint database are also needed in the Bluetooth-based methods.

Another popular method is Pedestrian Dead Reckoning (PDR), a localization method for calculating current position by referencing previously known location, speed, and facing direction [14]. Nowadays, almost everyone carries a smartphone and most of the smartphones are equipped with a wide array of sensors including a gyroscope, accelerometer and magnetometer which made it possible to estimate the facing direction and walking speed. However, PDR needs a starting point to initialize its position update process and also suffers from an error accumulation problem [15]. In addition, the three categories of positioning methods, Wi-Fi, Bluetooth and PDR, have insufficient localization or orientation accuracy for AR-based navigation.

As the price of camera sensors has been declining in recent years, vision-based positioning methods are becoming more and more popular. There are two main categories in a visual-based positioning and navigation system: marker-based and marker-free systems. Marker-based indoor positioning and navigation system works by placing a marker that can be recognized by the system to get the current position. The marker could be anything as long as the system could recognize it. One example of such a system is from research conducted by Sato [16]. The system proposed by Sato works by placing various AR markers either on the floor or wall all over the building. Each marker is a unique marker and localization can be done by scanning the marker. To overcome the limits on the number of distinct markers, Chawathe proposed to use a sequence of recently seen markers to determine locations [17]. Mulloni et al. demonstrated that marker-based navigation is easier to use than conventional mobile digital maps [18].

Romli et al. proposed an AR navigation for the smart campus [19]. The developed system used AR markers to guide users inside the library by providing the right direction and information instantly. Hartmann et al. proposed an indoor 3D position tracking system by integrating an Inertial Measurement Unit (IMU) and marker-based video tracking system [20]. The 3D position, velocity and attitude are calculated from IMU measurements and aided by using position corrections from the marker-based video tracking system. Koch et al. proposed to use the indoor natural markers already available on site, such as exit signs, fire extinguisher location signs, and appliance labels for supporting AR-based navigation and maintenance instructions [21]. Overall, the advantage of a marker-based navigation system is that since the marker is placed either systematically or at the known position, the system could easily perform localization. The disadvantage of the marker-based navigation systems is that the preparation and maintenance of markers are required. In some cases, real-time localization is not possible because the localization is dependent on the visibility of markers.

Another category is the marker-free vision-based positioning and navigation system. In contrast to the marker-based positioning and navigation system mentioned previously, the marker-free approach does not require setting up the marker in the building. This

simplifies the complexity of preparation or maintenance when the positioning and navigation system is deployed in a large-scale building. Several successful examples can be found for outdoor navigation, and these systems prefer to conduct the positioning and navigation in a pre-prepared street view image database. Robertson et al. developed a system using a database of views of building facades to determine the pose of a query view provided by the user at an unknown position [22]. Zamir et al. used 100,000 Google Maps Street View images as the map and developed a voting-based positioning method, in which the positioning result is supported by the confidence of localization in the neighboring area [23]. Kim et al. addressed the problem of discovering features that are useful for recognizing a place depicted in a query image by using a large database of geotagged images at a city-scale [24].

Torii et al. formulated image-based localization problem as a regression on an image graph with images as nodes and edges connecting nearby panoramic Google Street View images [25]. In their research, the query image is also a panoramic image with a 360-degree view angle. Sadeghi et al. proposed to find the two best matches of the query view from the Google Street View database and detect the common features among the best matches and the query view [26]. The localization was considered as a two-stage problem: estimating the world coordinate of the common feature and then calculating the fine location of the query. Yu et al. built a coarse-to-fine positioning system, namely a topological place recognition process and then a metric pose estimation by local bundle adjustment [27]. In order to avoid discrete localization problems in the Google Street View based localization, Yu et al. further proposed to virtually synthesize the augmented Street View data and render a smooth and metric localization [28]. However, there is not a pre-prepared geotagged image database for most of the indoor environments.

In addition, there are a few discussions for AR-based navigation. Mulloni et al. investigated user experiences when using AR as a new aid to navigation [29]. Their results suggested that navigation support will be most needed in the proximity of road intersections. Katz et al. proposed a stereo-vision-based navigation system for the visually impaired [30]. The proposed system can detect the important geo-located objects, and guide the user to his or her desired destination through spatialized semantic audio rendering. Hile et al. proposed a landmark-based navigation system for smartphones, which is the pioneer for AR-based pedestrian navigation [31]. In addition, Hile et al. pointed out that it is necessary to reduce the positioning error and provides accurate camera pose information for the more realistic visualization of arrows in the navigation [32].

The first contribution of this paper is to propose a marker-free system for AR-based indoor navigation. The Simultaneous Localization and Mapping (SLAM) method is a famous vision-based positioning method. It can accurately estimate the position and orientation of the vision sensor [33,34]. However, SLAM is still a positioning method, it cannot provide the navigation information directly for users. This research proposes an AR navigation system using SLAM to create a point cloud map which then will be integrated with a floor map to create a hybrid map containing a point cloud map of the indoor environment and navigation information on the floor map such as the position of doors and the name of rooms. The proposed AR navigation system performs positioning and navigation in this proposed hybrid map.

The second contribution is to propose an error correction method to improve the accuracy of the estimated camera orientation which is needed to visualize the augmented navigation information on the real scene correctly. It is observed in the feasibility step of this research that the estimated camera orientation from SLAM is not always correct. The orientation error is the reason for the incorrect navigation. To correct orientation error, a deep learning based real-time object detector You Only Look Once, or YOLO, developed by Redmon et al. [35] is used to detect the static objects, such as the door of a room. The error correction is conducted by minimizing the position difference between the detected doors from the image and the projected doors from the hybrid map.

The conception of using SLAM and floor map for AR-based navigation has been presented in our previous paper [36]. This paper extends the initial results and newly proposes the orientation error correction to further improve the performance of the navigation system. This paper is a summary of the undergraduate research of the first author [37].

2. Proposed AR-Based Navigation System

Oriented Brief (ORB)-SLAM is a real-time SLAM library [33] for Monocular, Stereo and RGB-D cameras that computes the camera trajectory and a sparse 3D reconstruction (in the stereo and RGB-D case with true scale). It can detect loops and relocalize the camera in real time. ORB-SLAM provides both SLAM Mode and Localization Mode [8]. The proposed AR-based navigation system is built based on the ORB-SLAM. In addition, the hybrid map and orientation error correction function is proposed to realize the accurate AR-based navigation.

Figure 1 shows the overview of the proposed AR-based navigation system. The proposed system consists of two parts. The first part is mapping, and the other part is positioning and navigation. In this research, an RGB-D camera is used to collect environment data. ORB-SLAM is then used to generate a point cloud map. The generated point cloud map is later manually integrated with a floor map and finally forms a hybrid map. The positioning and navigation are performed on the hybrid map. The proposed system employs ORB-SLAM for the initial positioning. After that, the developed orientation error method optimizes the initial positioning result by referring to the position of the detected objects provided by a deep neural network-based object detection model. Finally, the navigation information is augmented in the real scene image to realize the AR-based navigation.
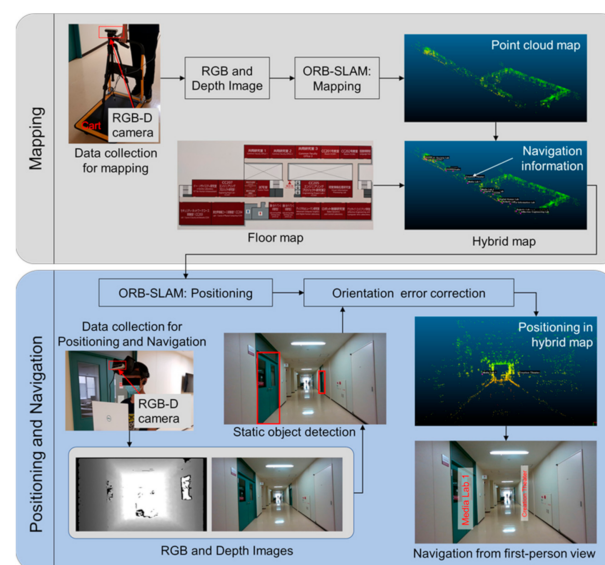

Figure 1. Overview of the proposed indoor navigation system.

2.1. Generation of Hybrid Map

Generation of the hybrid map starts with generating a point cloud map. This research adopts RGB-D camera for sensing in the navigation system. RGB and depth image sequence is collected to generate the point cloud map. By using depth images, accurate depth information can be obtained. The accurate depth information can solve the uncertain scaling problem that is common in monocular camera-based SLAM.

In this research, ORB-SLAM mode is employed to create the point cloud map. Basically, ORB-SLAM consists of three threads: tracking, mapping and loop closing. In the tracking thread, the ORB features are firstly extracted from the RGB and depth images, the initial pose estimation is performed according to an initialized pose through global relocation, or the pose in the previous frame based on constant velocity motion model. Then, the initial

pose is optimized by tracking the reconstructed local map. After that, a new keyframe is decided. The mapping thread is mainly responsible for the local map construction. Inserting key frames, selection of map points, removing redundant key frames, a series of processing are conducted to build the local point cloud map around of the camera pose. In the loop closing thread, if a loop is detected, a similar transformation is calculated, and the accumulated drift is estimated in the loop. Finally, posture graph optimization is performed on similar constraints to achieve global consistency. Figure 2 shows the constructed point cloud map (by ORB-SLAM) along a corridor in a large-scale building.
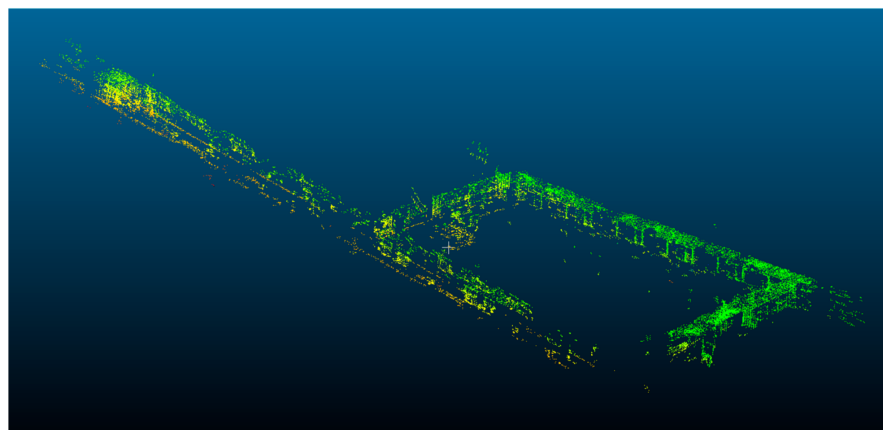

Figure 2. Point cloud map of corridor in a large-scale building.

A point cloud map only contains the feature points of the indoor environment. The navigation information, such as the door position and room information, is not included in the point cloud map. The floor map, as shown in Figure 3, represents the layout of each floor in a building. All the necessary information for navigation has been indicated in the floor map. This research proposes to extract the navigation information from the floor map and augment the navigation information in the point cloud map. The proposed AR-based navigation system will show the name of each room at the area of door from the first-person view in the navigation. Therefore, the door position and the name of each room are extracted from the floor map.

Figure 3. Floor map in a large-scale building.

Firstly, the floor map and the point cloud map are aligned. Then, the door position and the name of each room are labeled in the point cloud map by referring to the floor map. The point at the center of each door is labeled as the door position. In addition, the name of the corresponding room is an annotation of the point as shown in Figure 4. In this way, if the annotated point appears in the camera, the name of the room can be visualized immediately. Currently, this extraction of the navigation information from the floor map is conducted manually. In the future, a fully automatic extraction and labeling

algorithm will be developed. Figure 5 visualizes the developed hybrid map. The hybrid map obtained after integrating the point cloud map with the floor map contains a feature points coordinate that can be used for positioning, and a door coordinate and room name that can be used for navigation.
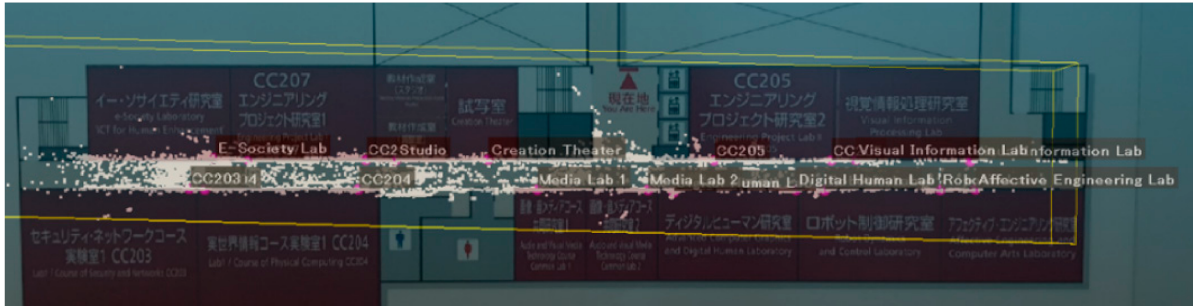

Figure 4. Extraction of the navigation information from floor map. The purple points are labeled as the center of each door. The white color text annotated on the purple point indicates the name of the corresponding room.
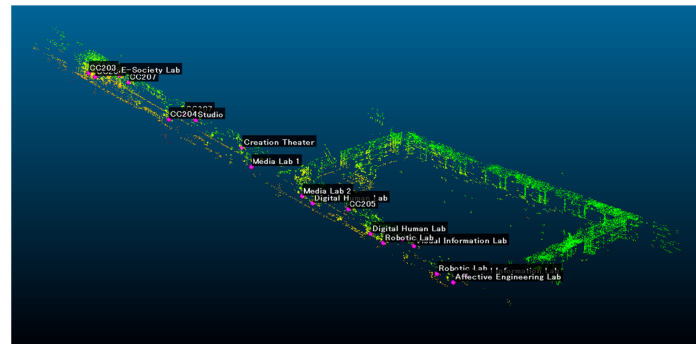

Figure 5. Hybrid map containing point cloud data and navigation information.

2.2. Positioning and Navigation

The process of positioning is also conducted by ORB-SLAM. Different from point cloud map generation, only tracking part is used while the local mapping and loop detection are not running during positioning. In this step, the current position coordinate in X, Y, and Z is obtained along with orientation data in quaternion (q_x, q_y, q_z, q_w). Further conversion from quaternion to rotation matrix can be expressed as shown in Equation (1) below.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} 1 - 2q_y^2 - 2q_z^2 & 2q_xq_y + 2q_zq_w & 2q_xq_z - 2q_yq_w \\ 2q_xq_y - 2q_zq_w & 1 - 2q_x^2 - 2q_z^2 & 2q_yq_z + 2q_xq_w \\ 2q_xq_z + 2q_yq_w & 2q_yq_z - 2q_xq_w & 1 - 2q_x^2 - 2q_y^2 \end{bmatrix} \quad (1)$$

where R denotes rotation matrix, q_x, q_y, q_z, and q_w are quaternion representing spatial orientations and rotations. The rotation matrix R obtained in this step can later be used for navigation.

To visualize the room name at the area of door in the real scene image, the pixel coordinate of the purple point (from the hybrid map) needs to be calculated in Equation (2).

$$S \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_X & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2)$$

The result of Equation (2) is the pixel coordinate of the purple point denoted as u and v, where u is the pixel coordinate in X axis and v is the pixel coordinate in Y axis. The camera parameters are denoted as f_X, f_y, C_x, and C_y. The f_X and f_y parameters are the focus

length of the camera. The C_x and C_y parameters are the center of the camera. The values of f_X, f_y, C_x, and C_y can be obtained through a camera calibration process. r_{11}–r_{33}, are the parameters in the rotation matrix that are obtained by converting quaternion to rotation matrix that was shown in Equation (1). $[t_1, t_2, t_3]$ is the translation of the camera from point of origin. t_1 is in X axis, t_2 is in Y axis, and t_3 is in Z axis. $[t_1, t_2, t_3]$ is the position coordinate of the camera which is obtained from the positioning step using ORB-SLAM. In the last matrix, X, Y, and Z are the coordinates of the purple point in the hybrid map. Figure 6 demonstrates the process of the projection. Figure 6a is the real scene image, Figure 6b is the hybrid map viewed from the camera position of Figure 6a, Figure 6c is the augmented image with navigation information. From the augmented navigation information, users can know the name from each room without approaching the room. In fact, more details of the room, such as the functionality and facility in the room, can be visualized on the image. This research only uses the name of the room to demonstrate the proposed system.
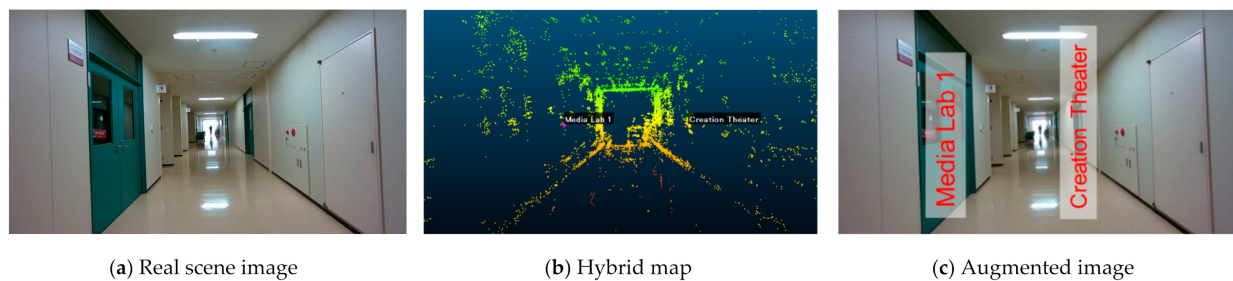

(**a**) Real scene image   (**b**) Hybrid map   (**c**) Augmented image

Figure 6. Visualization of navigation information in real scene image; (**a**) is the real scene image, (**b**) is the hybrid map viewed from the camera position of (**a**), (**c**) is the augmented image with navigation information.

In addition to the room name, the proposed system also provides the direction arrow in the navigation as shown in Figure 7. Currently, there are four navigation arrow modes, forward, backward, left, and right. If the target room is not within the range where the navigation label would be augmented, which in the current configuration is 2.7 m, the arrow would be either forward or backward depending on the orientation of the camera. If the target landmark (e.g., the door of the room) is within the range where the navigation label is augmented on the screen, the arrow would show either left or right. The left or right arrow would only show if the door of the target room can be visually seen on the image sequence. This relative position is calculated based on the position of door and camera positioning and orientation.

Figure 7. Navigation with direction arrow.

2.3. Orientation Error Correction in Navigation

It is observed in this step that the orientation value obtained from ORB-SLAM is not perfectly correct. This error, especially for the error at yaw direction, will lead to placing the augmented navigation information at a wrong place, for example, outside of the door area in the image. This paper proposes to detect the door of the room from images and match the detected doors with the doors (labeled as purple points) in the hybrid map to correct the orientation error.

In this research, a deep learning based real-time object detector You Only Look Once, (YOLO), developed by Redmon et al. [35], is used to detect the door. The model used in the object detector was newly trained using manually labeled images with about 830 door instances. The detector has about 96% detection rate in the test sequence. The output generated from the detector are the pixel coordinate of the bounding box of each door as shown in Figure 8. The center of the detected door can be calculated from the coordinate of the bounding box directly. In addition, the center of each door in the camera space can also be obtained from the hybrid map based on Equation (2). If the estimated orientation from ORB-SLAM is correct, the purple points should be located at the center of the detected doors. To correct the orientation error and visualize the room name on the door area correctly, this research proposes to estimate the orientation error by minimizing the difference between the center of the detected door and the projected center. In this research, the orientation error at yaw direction has a significant impact on the correctness of visualization. Therefore, the error correction focuses on the yaw direction. The proposed orientation error correction method can be expressed by Equations (3)–(5).

$$\theta_{min} = \underset{\theta \in [-n\Delta_\theta, -(n-1)\Delta_\theta, \ldots, n\Delta_\theta]}{\arg\min} \{h(\theta)\}, h(\theta) = \frac{1}{n} \sum_{i=1}^{n} |u(\theta) - u_{detected}| \tag{3}$$

$$S(\theta) \begin{bmatrix} u(\theta) \\ v(\theta) \\ 1 \end{bmatrix} = \begin{bmatrix} f_X & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} & & t_1 \\ g(\theta) \cdot R & & t_2 \\ & & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4}$$

$$g(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{5}$$

where θ_{min} is the orientation error at yaw direction and $h(\theta)$ is a function for finding the average distance between the projected center of the door and the center of the detected door. $u(\theta)$ denotes the pixel coordinate of the projected door center at the horizontal direction of the image. $u(\theta)$ can be calculated from Equation (4). The parameters on the right side of Equation (4) are the same as the parameters in Equation (2). $u_{detected}$ is the pixel coordinate of the detected door center at the horizontal direction of the image. $g(\theta)$ is the rotation matrix caused by the yaw correction angle. In this research, n and Δ_θ are empirically set as 20 and 0.005. After the estimation of orientation error, the new position (u, v) of the navigation information can be calculated based on Equation (6):

$$S \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_X & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta_{min}) & 0 & \sin(\theta_{min}) & t_1 \\ 0 & 1 & 0 & R & t_2 \\ -\sin(\theta_{min}) & 0 & \cos(\theta_{min}) & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{6}$$

In this research, only a door is used as the object for orientation error correction. It is also possible to detect other kinds of objects to support the error correction.
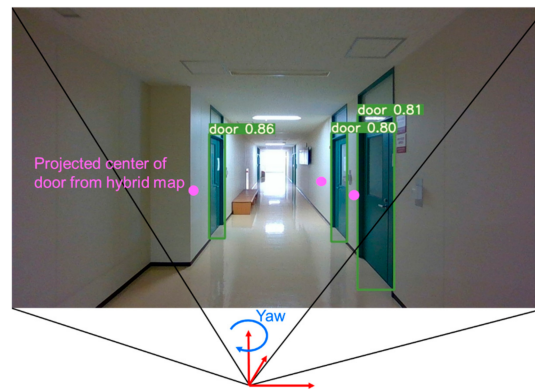
Figure 8. Orientation error correction with the aid of door detection. The green rectangle is the bounding box of the detected door. The purple point is the projected center of each door from the hybrid map. The error correction is conducted at the yaw direction by minimizing the difference between the center of the detected door and the projected center.

3. Experiments

3.1. Experiment Setup

In this research, an Intel Real Sense D455 RGB-D camera was used. In the experiment, RGB image and depth image are captured synchronously; the resolution of RGB and depth image is 1280 × 720 pixels. The images are recorded at 30 frames per second (FPS). As shown in Figure 9a, the camera attached to a tripod was setup on a cart in the mapping step. The camera on a cart ensures a stable image sequence to provide the best point cloud map generation quality. In addition, data used for position and navigation are collected by placing the camera on the forehead as shown in Figure 9b. This can simulate actual usage with unstable motion when the user is moving and wearing smart glasses.
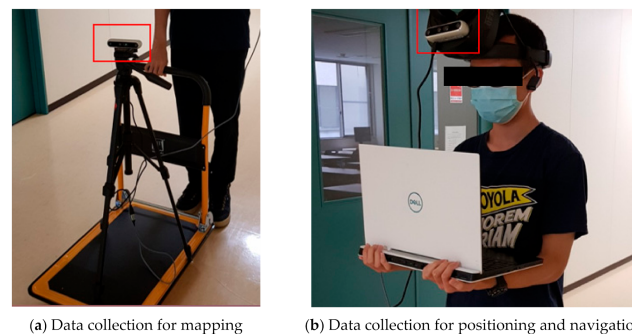

(**a**) Data collection for mapping   (**b**) Data collection for positioning and navigation

Figure 9. Data collection for mapping step (**a**), positioning and navigation step (**b**).

3.2. Experiment Result of Room Information Visualization

To evaluate the quality of the navigation, this research conducted a series of experiments in the corridor of a largescale building. Table 1 shows the comparison between different methods with or without orientation error correction. The value offset refers to the pixel offset from the center of the visualized label to the centerline of the door in image. In addition, the evaluation also considers in how many instances the center of the navigation information label is placed in the area of the door. As shown in Table 1, orientation error correction significantly improved the performance of the system. The average offset is reduced to 15 pixels by using the method with orientation error correction. Moreover, the room name can be correctly visualized within the door area in 81.82% of frames if the orientation error correction is used.

Table 1. Comparison of visualization correctness in navigation.

Method	Average Offset (Pixel)	Navigation Label Visualized within Door Area
No correction (baseline)	43	42.45%
With orientation error correction	15	81.82%

Several experiment results are demonstrated in Figure 10; the orientation value obtained from ORB-SLAM in this specific image is already correct. There is not so much difference between the visualization results generated by methods without correction (Figure 10a) and with orientation error correction (Figure 10b).

(**a**) Visualization result generated by the method without correction

(**b**) Visualization result generated by the method with orientation error correction

Figure 10. Visualization of navigation information is correct in both the method without correction (**a**) and the method with orientation error correction (**b**).

In Figure 11, the orientation obtained from ORB-SLAM in this image is incorrect. This error causes the navigation label augmented in the baseline method to also be incorrect (in Figure 11a, navigation labels are augmented outside the door area). In this scenario, the orientation error correction method works and performs better than the baseline method. The room name is correctly visualized at the door area (Figure 11b).

(**a**) Visualization result generated by the method without correction

(**b**) Visualization result generated by the method with orientation correction

Figure 11. Visualization of navigation information is incorrect in the method without correction (**a**), but correct in the method with orientation error correction (**b**).

3.3. Experiment Result of Navigation Arrow Visualization

In addition to the room name, the proposed system also can visualize a direction arrow for navigation. As for navigation arrow augmentation, the system achieved 99% accuracy in the test. For example, the room called "Studio" was set as a target for navigation. On the left, the studio room is not within augmentation range, therefore the forward arrow is augmented on the image. In the middle of Figure 12, the studio room is within the augmentation range and is estimated to be on the right side of the camera, therefore, the

right navigation arrow is augmented. On the right of Figure 12 when studio room is immediately outside the image, the backward navigation arrow is augmented.

Figure 12. Different augmented navigation arrows visualized in the navigation to a destination "Studio". **Left**: the destination is in front but not within the augmentation range. **Middle**: the destination is on the right side of the user. **Right**: The user has passed the destination.

4. Discussion

The proposed orientation error correction can improve the correctness of the visualization of the navigation information on the real scene image. The orientation error correction function works when the reference objects (the door of the room in this paper) appear in the image. The proposed system utilizes ORB-SLAM for mapping and positioning. ORB-SLAM focuses on building globally consistent maps for reliable and long-term localization in a wide range of environments as demonstrated in the experiments of the original ORB-SLAM paper [33]. The original paper of ORB-SLAM provides the evaluation for the accuracy of ORB-SLAM for RGB-D sequence. In our experiment, the whole floor of the laboratory's building is successfully mapped. The area of the floor is about 100 by 30 m, so it is reasonable to say that the current system can work in such a scale area.

However, the proposed system has some limitations. The process of extracting navigation information is a manual process that requires an accurately scaled floor map. Automatic navigation extraction from the floor map is one significant area for improving the system. Furthermore, for a large-scale area, position initialization could take a significant amount of time. For a seamless AR experience, another positioning method (e.g., Wi-Fi-based positioning) could be one of the possible solutions to speed up the position initialization process.

The proposed AR-based navigation system is demonstrated in a classroom building scenario and it can be applied in other environments, such as shopping malls and museums. In addition, if this standalone system is connected to the internet, and the owner of the system shares the tour with others, the proposed system will become a virtual tour tool. The proposed system can visualize both the real scene and the augmented information, which will improve the experience for remote users. Moreover, the audio augmented reality can be added to the proposed system devoted to visually impaired people.

5. Conclusions

This research proposed an AR-based indoor navigation system. The proposed system adopts the RGB-D camera to observe the surrounding environment and relies on ORB-SLAM for mapping and positioning. In addition, this research proposed a hybrid map created from a combination of a point cloud map generated by ORB-SLAM and a floor map. The proposed hybrid map can be used for both navigation and positioning. Furthermore, to improve the correctness of navigation, an orientation error correction method was proposed. The proposed method corrects orientation error by finding the optimum rotation matrix that results in the lowest average offset from the position of projected objects to the center of detected objects. The proposed orientation error correction method achieved an average offset of about 15 pixels and 81.82% of navigation labels are augmented within the boundary of door. The navigation arrow augmentation achieved a very high accuracy of 99%. The

experimental results indicated that the current SLAM technology still needs improvement for providing correct navigation information. This study presents an example of using the object detection-aided orientation error estimation method to correct the orientation error and correctly visualize navigation information. The proposed AR-based navigation system can provide first-person view navigation with satisfactory performance. In the future, the proposed system will be improved for adaptation to the shopping mall environment.

References

1. Ohm, C.; Bienk, S.; Kattenbeck, M.; Ludwig, B.; Müller, M. Towards Interfaces of Mobile Pedestrian Navigation Systems Adapted to the User's Orientation Skills. *Pervasive Mob. Comput.* **2016**, *26*, 121–134. [CrossRef]
2. Bolton, A.; Burnett, G.; Large, D.R. An Investigation of Augmented Reality Presentations of Landmark-based Navigation using a Head-up Display. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Nottingham, UK, 1–3 September 2015.
3. Huang, B.C.; Hsu, J.; Chu, E.T.H.; Wu, H.M. ARBIN: Augmented Reality Based Indoor Navigation System. *Sensors* **2020**, *20*, 5890. [CrossRef] [PubMed]
4. Wang, H.; Gu, Y.; Kamijo, S. Pedestrian Positioning in Urban City with the Aid of Google Maps Street View. In Proceedings of the 15th IAPR International Conference on Machine Vision Applications, Nagoya, Japan, 8–12 May 2017.
5. Rehman, U.; Cao, S. Augmented-reality-based Indoor Navigation: A Comparative Analysis of Handheld Devices versus Google Glass. *IEEE Trans. Hum. Mach. Syst.* **2016**, *47*, 140–151. [CrossRef]
6. Laoudias, C.; Moreira, A.; Kim, S.; Lee, S.; Wirola, L.; Fischione, C. A Survey of Enabling Technologies for Network Localization, Tracking, and Navigation. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 3607–3644. [CrossRef]
7. Vaupel, T.; Seitz, J.; Kiefer, F.; Haimerl, S.; Thielecke, J. Wi-Fi Positioning: System Considerations and Device Calibration. In Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation, Zurich, Switzerland, 15–17 September 2010.
8. Pei, L.; Liu, J.; Guinness, R.; Chen, Y.; Kuusniemi, H.; Chen, R. Using LS-SVM based Motion Recognition for Smartphone Indoor Wireless Positioning. *Sensors* **2012**, *5*, 6155–6175. [CrossRef] [PubMed]
9. Wang, X.; Gao, L.; Mao, S.; Pandey, S. CSI-based Fingerprinting for Indoor Localization: A Deep Learning Approach. *IEEE Trans. Veh. Technol.* **2017**, *66*, 763–776. [CrossRef]
10. Huang, Y.; Hsu, L.T.; Gu, Y.; Wang, H.; Kamijo, S. Database Calibration for Outdoor Wi-Fi Positioning System. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2016**, *99*, 1683–1690. [CrossRef]
11. Oguejiofor, O.S.; Aniedu, A.N.; Ejiofor, H.C.; Okolibe, A.U. Trilateration based Localization Algorithm for Wireless Sensor Network. *Int. J. Sci. Mod. Eng.* **2013**, *10*, 2319–6386.
12. Mok, E.; Retscher, G. Location Determination using Wi-Fi Fingerprinting versus Wi-Fi Trilateration. *J. Locat. Based Serv.* **2017**, *2*, 145–159.
13. Faragher, R.; Harle, R. An Analysis of the Accuracy of Bluetooth Low Energy for Indoor Positioning Applications. In Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation, Tampa, FL, USA, 8–12 September 2014.
14. Hsu, L.T.; Gu, Y.; Huang, Y.; Kamijo, S. Urban Pedestrian Navigation using Smartphone-based Dead Reckoning and 3-D Map-aided GNSS. *IEEE Sens. J.* **2016**, *5*, 1281–1293. [CrossRef]
15. Gu, Y.; Li, D.; Kamiya, Y.; Kamijo, S. Integration of Positioning and Activity Context Information for Lifelog in Urban City Area. *J. Inst. Navig.* **2020**, *67*, 163–179. [CrossRef]
16. Sato, F. Indoor Navigation System based on Augmented Reality Markers. In Proceedings of the International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Torino, Italy, 10–12 July 2017.
17. Chawathe, S.S. Marker-based Localizing for Indoor Navigation. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Bellevue, WA, USA, 30 September–3 October 2007.
18. Mulloni, A.; Wagner, D.; Barakonyi, I.; Schmalstieg, D. Indoor Positioning and Navigation with Camera Phones. *IEEE Pervasive Comput.* **2009**, *8*, 22–31. [CrossRef]

19. Romli, R.; Razali, A.F.; Ghazali, N.H.; Hanin, N.A.; Ibrahim, S.Z. Mobile Augmented Reality (AR) Marker-based for Indoor Library Navigation. In Proceedings of the 1st International Symposium on Engineering and Technology 2019, Perlis, Malaysia, 23 December 2019.

20. Hartmann, B.; Link, N.; Trommer, G.F. Indoor 3D Position Estimation using Low-cost Inertial Sensors and Marker-based Video-Tracking. In Proceedings of the IEEE/ION PLANS 2010, Indian Wells, CA, USA, 4–6 May 2010.

21. Koch, C.; Neges, M.; König, M.; Abramovici, M. Natural Markers for Augmented Reality-based Indoor Navigation and Facility Maintenance. *Autom. Constr.* **2014**, *48*, 18–30. [CrossRef]

22. Robertson, D.P.; Cipolla, R. An Image-Based System for Urban Navigation. In Proceedings of the 2004 British Machine Vision Conference, Kingston, UK, 7–9 September 2004.

23. Zamir, A.R.; Shah, M. Accurate Image Localization based on Google Maps Street View. In Proceedings of the 2010 European Conference on Computer Vision, Crete, Greece, 5–11 September 2010.

24. Kim, H.J.; Dunn, E.; Frahm, J.M. Predicting Good Features for Image Geo-Localization Using Per-Bundle VLAD. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.

25. Torii, A.; Sivic, J.; Pajdla, T. Visual Localization by Linear Combination of Image Descriptors. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops, Barcelona, Spain, 6–13 November 2011.

26. Sadeghi, H.; Valaee, S.; Shirani, S. 2DTriPnP: A Robust Two-Dimensional Method for Fine Visual Localization Using Google Streetview Database. *IEEE Trans. Veh. Technol.* **2017**, *66*, 4678–4690. [CrossRef]

27. Yu, L.; Joly, C.; Bresson, G.; Moutarde, F. Monocular Urban Localization using Street View. In Proceedings of the 2016 IEEE 14th International Conference on Control Automation, Robotics and Vision, Phuket, Thailand, 13–15 November 2016.

28. Yu, L.; Joly, C.; Bresson, G.; Moutarde, F. Improving Robustness of Monocular Urban Localization using Augmented Street View. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems, Rio de Janeiro, Brazil, 1–4 November 2016.

29. Mulloni, A.; Seichter, H.; Schmalstieg, D. User Experiences with Augmented Reality aided Navigation on Phones. In Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, Switzerland, 26–29 October 2011.

30. Katz, B.F.; Kammoun, S.; Parseihian, G.; Gutierrez, O.; Brilhault, A.; Auvray, M.; Truillet, P.; Denis, M.; Thorpe, S.; Jouffrais, C. NAVIG: Augmented Reality Guidance System for the Visually Impaired. *Virtual Real.* **2012**, *16*, 253–269. [CrossRef]

31. Hile, H.; Vedantham, R.; Cuellar, G.; Liu, A.; Gelfand, N.; Grzeszczuk, R.; Borriello, G. Landmark-based Pedestrian Navigation from Collections of Geotagged Photos. In Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia, Umea, Sweden, 3–5 December 2008.

32. Hile, H.; Grzeszczuk, R.; Liu, A.; Vedantham, R.; Košecka, J.; Borriello, G. Landmark-based Pedestrian Navigation with Enhanced Spatial Reasoning. In Proceedings of the 2009 International Conference on Pervasive Computing, Nara, Japan, 11–14 May 2009.

33. Mur-Artal, R.; Tardós, J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [CrossRef]

34. Engel, J.; Schöps, T.; Cremers, D. LSD-SLAM: Large-scale Direct Monocular SLAM. In Proceedings of the 2014 European Conference on Computer Vision (ECCV2014), Zurich, Switzerland, 6–12 September 2014.

35. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Vegas, NV, USA, 27–30 June 2016.

36. Gu, Y.; Chidsin, W.; Goncharenko, I. AR-based Navigation Using Hybrid Map. In Proceedings of the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies, Nara, Japan, 9–11 March 2021.

37. Chidsin, W. Augmented Reality-based Indoor Navigation Using SLAM and Hybrid Map Information. Undergraduate Thesis, Ritsumeikan University, Kyoto, Japan, 1 February 2021.