Position Estimation Method based on Object Detection of 3D Reconstructed Map

Sukwoo Jung
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, South Korea
swjung@keti.re.kr

Youn-Sung Lee
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, South Korea
yslee@keti.re.kr

KyungTaek Lee
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, South Korea
ktechlee@keti.re.kr

***Abstract*— Recent advancements in sensor-based position estimation have gained significant attention due to their widespread applicability across diverse domains, including augmented reality (AR), virtual reality (VR), robotics, and beyond. While existing state-of-the-art indoor tracking methods predominantly rely on technologies like Wi-Fi or Bluetooth, this research presents an innovative departure from conventional approaches. Rather than solely depending on images or Inertial Measurement Unit (IMU) sensors for position estimation, we introduce a novel method that uses pre-reconstructed 3D maps to drastically enhance the accuracy of location tracking. To accomplish this, we harness high-precision indoor maps obtained through the Leica RTC360 laser scanner, which serves as the foundation for training our system. This, in turn, enables us to proficiently track objects using data from camera sensors and precisely calculate the positional coordinates within indoor environments.**

Keywords—3D reconstruction, pose estimation, navigation, feature matching, object detection

I. INTRODUCTION

The evolution of modern technology has ushered in a new era where position tracking plays a pivotal role across a spectrum of applications. It has become indispensable in fields such as Virtual Reality (VR), Augmented Reality (AR), gaming, security systems, and robotics. In today's landscape, widely adopted position estimation systems primarily come in the form of smartphone applications [1] or standalone devices. These systems heavily rely on mapping services and satellite localization for their functionality.

While the Global Positioning System (GPS) is the go-to choice for outdoor scenarios, it falls short in reliability when deployed indoors. On the other hand, cutting-edge indoor localization methods utilize the devices known as beacons. These strategically placed beacons emit unique identifiers and predominantly operate via Bluetooth or other radio frequency (RF) technologies, functioning effectively over short distances. However, their implementation necessitates the deployment of a complex infrastructure. Alternatively, WiFi based localization demands the establishment of a dense network of access points, resulting in substantial maintenance costs, while offering only limited positional accuracy.

Various studies have explored methods for position tracking, employing either single or multiple sensors [2] such as Inertial Measurement Units (IMUs) and Bluetooth, or approaches based on the recognition of feature points [3] within images, including Simultaneous Localization and Mapping (SLAM). However, these methods often have limitations in terms of accuracy and real-time performance.

In this paper, we present a method that combines 3D spatial reconstruction information with object recognition technique to enhance the accuracy and efficiency of estimating the position.

II. PROPOSED METHOD

3D spatial reconstruction refers to the process of reconstructing the 3D shape and 2D texture of the objects or environments in 3D space, primarily from 2D images or videos [4]. This process entails the precise alignment of data from various of sensors. In our study, we utilized pre-reconstructed 3D map information as a fundamental component of our position tracking algorithm.

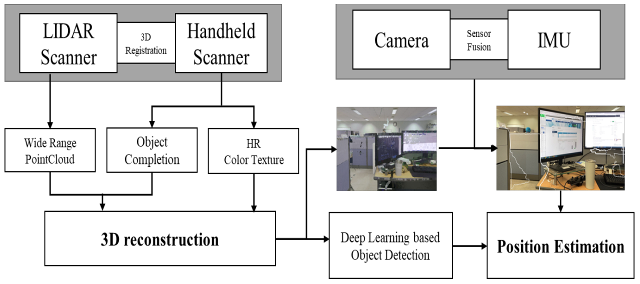

Fig. 1. Position estimation using 3D reconstructed map workflow

The overall procedure of our proposed method is vividly illustrated in Figure 1. To begin the process, we obtained point cloud data by employing the Leica RTC360 scanner for 3D spatial reconstruction. These diverse multiple sets of 3D data were merged to create a unified reconstructed map [5]. From this map, we extracted various types of features points and generated templates for matching.

Next, we collected data from both cameras and the Inertial Measurement Unit (IMU) to estimate the pose of the sensor. The fusion of these two sets of sensor data enhanced the precision of location tracking [6]. Subsequently, we applied SLAM to video obtained from the camera, enabling the simultaneous acquisition of the sensor's position and the 3D information of the surrounding environment.

Finally, by matching the surrounding environment with the pre-reconstructed 3D map, we determine the optimal position, allowing us to ascertain the location of recognized objects.

Throughout this process, we adeptly harnessed both 3D data and image data for template matching, effectively estimating the objects position within the 3D map. Furthermore, this enables the estimation of the three-dimensional coordinates of both the user and the objects, opening up the possibilities for a wide use of applications.

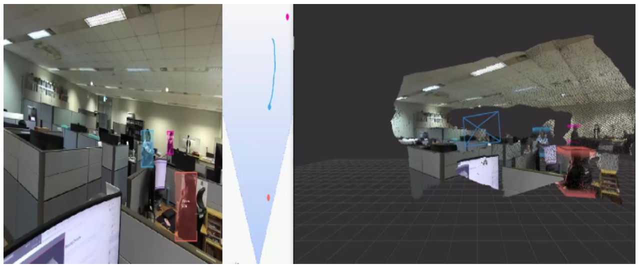

Fig. 2. The position estimation result of the proposed method.

The proposed method's implementation was carried out using the C++ programming language. Testing was performed to validate the effectiveness of the proposed position tracking method. The results of this testing are depicted in Figure 2. In this illustration, we can see the object of interest being successfully tracked and visualized using the Bird's Eye View (BEV) trasnformation. We have plans to conduct further experiments in the future to compare our system against existing algorithms in a multitude of diverse environments is essential. It is important to evaluate and analyze the performance of our system thoroughly, especially concerning its accuracy and real-time capabilities.

While the system generally performs reasonably well under typical conditions, we did observe some noteworthy limitations. Particularly, issues arise when the camera's Field of View (FoV) becomes obstructed by dynamic objects or when the background experiences substantial alterations compared to the previously reconstructed 3D map information. In such cases, the background recognition of the prosposed algorithm does not function correctly.

It's important to understand that position estimation involves dealing with various challenges. These include calibrating the multiple sensors, reducing noise or errors, and dealing with drift of sensor over time. In addition, comparing accuracy in position estimation is difficult because measuring ground truth is hard. By addressing these challenges, we aim to push the boundaries of position estimation technology, ultimately contributing to its ongoing evolution and improvement.

III. CONCLUSION

In summation, this paper has introduced a position tracking method that ingeniously merges 3D reconstruction information with object detection technology. This method can enhances the user experience across diverse applications and is poised to make contributions to future technological advancements.

As we draw the curtains on this endeavor, it is of paramount importance to underscore the significance of perpetual research and constant innovation in the sphere of 3D map based position estimation. We emphasize the critical role of ongoing research and innovation in position estimation algorithms, which must continually evolve to meet the burgeoning demands of evolving technologies and applications.

It's essential to recognize the limitations when dealing with significant occlusions in images or substantial changes in previously reconstructed maps, which can potentially affect the object recognition accuracy. These challenges can be adressed by harnessing dynamic object recognition technology to distinguish moving objects from the background and by deploying SLAM technology to perceive the environment and update the map dynamically. Therefore, to overcome these limitations and make progress in position estimation, we need to stay dedicated to ongoing research.

ACKNOWLEDGMENT

This research was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (MSIT) (No.2023-0-00970, Development of multi-image-based 3D virtual space and dynamic object reconstruction technology for manufacturing site support).

REFERENCES

[1] Gerstweiler, Georg, Emanuel Vonach, and Hannes Kaufmann. "HyMoTrack: A mobile AR navigation system for complex indoor environments." *Sensors* 16.1 (2015): 17.

[2] Bhorkar, Gaurav. "A survey of augmented reality navigation." arXiv preprint arXiv:1708.05006 (2017).

[3] Ashour, Ziad, Zohreh Shaghaghian, and Wei Yan. "BIMxAR: BIM-empowered augmented reality for learning architectural representations." *arXiv preprint arXiv:2204.03207* (2022).

[4] Sukwoo Jung, et al. "3D reconstruction using 3D registration-based ToF-stereo fusion." Sensors 22.21 (2022): 8369.

[5] Sukwoo Jung, et al. "Range image registration based on 2D synthetic images." Computer-Aided Design 94 (2018): 16-27.

[6] Sukwoo Jung, Seho Park, and KyungTaek Lee. "Pose tracking of moving sensor using monocular camera and IMU sensor." KSII Transactions on Internet and Information Systems (TIIS) 15.8 (2021): 3011-3024.