70 Image-Based Localization for Facilitating Construction Field Reporting on Mobile Devices

Youyi Feng and Mani Golparvar-Fard

Abstract

Current studies reveal the exceptional advantages of high-efficiency onsite information management for facilitating the design/building progress of construction projects. In particular, the prevailing methods of construction field reporting still primarily rely on manual on-site documentation of project information. Fortunately, state of the art computerized technologies provides solutions with great potential for boosting the efficiency of gathering and managing on-site information for field reporting. Providing on-demand access to such information in real-time requires an autonomous method for localizing and tracking (i.e. calculating the position and orientation) of a construction filed reporter on job site. This, in fact, will reduce both working time and efforts for providing on-demand access to project information. Mobile devices such as smartphones/tablets can be utilized to enable the on-site personnel to manage the project information in a portable fashion while adopting cloud technology for instant online access. In this paper, we proposed a method for on-site localization that can estimate and track the position and orientation of a hand-held device in a near real-time manner. The developed method is infrastructure-independent and marker-less. The proposed method mainly consists of mapping, localization and alignment modules. Initially, a video stream is acquired using the built-in camera of the mobile device scanning the target building. A 3D point cloud is then reconstructed from the acquired video data. Afterward, the localization algorithm outputs the location/orientation of the queried images using feature-based matching with the base 3D point cloud map generated earlier. Finally, global localization of frames is estimated by using the alignment parameters of the 3D point cloud with a Geo-referenced BIM model to transform the localized frames to the global reference. The proposed solution enables the field reporter to access, retrieve, save and edit the project information more efficiently on the construction site.

Keywords

Field reporting • Localization • SLAM • Point cloud • BIM

70.1 Introduction

Recent research efforts have been focused on improving information management on construction sites to provide on-demand access to project information and facilitate field reporting, whereas the prevailing methods of construction field reporting still primarily rely on written documents where the on-site field reporters need to manually record and update project information. During this process, field reporters have to carry large stacks of specifications and drawings with them which is usually inconvenient. Advances in computer vision techniques, as well as the hardware of mobile devices, actualized the idea of boosting the efficiency of gathering and managing project information on construction sites for facilitating the process of field reporting.

The fact that construction field reports are location-based makes it important to devise automated systems for localizing and tracking position/orientation of construction facilities/personnel on a job-site. Such system will reduce both working

Y. Feng (✉) · M. Golparvar-Fard
Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: feng41@illinois.edu

M. Golparvar-Fard
e-mail: mgolpar@illinois.edu

I. Mutis and T. Hartmann (eds.), *Advances in Informatics and Computing in Civil and Construction Engineering*,
https://doi.org/10.1007/978-3-030-00220-6_70

time and resources required for providing on-demand access to project information. With the rapid development of the computational power of portable handheld devices as well as computer vision-based techniques, mobile devices such as smartphones, tablets have created an unprecedented opportunity to enhance both the performance and efficiency of the current practice of construction field reporting especially while utilizing cloud-computing technology.

70.2 Background

Recent research efforts for improving accuracy and efficiency of construction fielding reporting have primarily focused on developing mobile Augmented Reality (AR) techniques to provide on-demand access to project information. However, to generate such a prompt and accurate field report, precisely tracking the position and orientation of onsite facilities as well as the field reporter himself on a construction job site is a crucial procedure [1, 2].

Prior research work reveals that the primarily applied localization approaches can be divided into four different categories based on the techniques used for estimating the pose: (i) sensor-based localization, where the user's position is tracked by using wireless local network (WLAN), GPS, and/or other sensors such as gyroscope, magnetometer sensors; (ii) fiduciary marker-based localization, where the user's location/orientation is determined by using predefined/preinstalled optical fiduciary markers; (iii) model-based localization, in this method, prior information (pre-built 3D point cloud of the physical world) is used to identify the user's relative location and orientation. (iv) Visual Simultaneous Localization and Mapping (Visual SLAM), algorithms for tracking and mapping visual features from images/videos to compute the user's location information.

Although the approaches stated above have the potential to provide fast location information to field personnel, they still have their drawbacks. Firstly, sensor-based methods, which uses RF-based (Radio Frequency) location tracking techniques, e.g. GPS and WLAN. They actually heavily rely on pre-installed infrastructures, such as wireless access points and GPS satellite receivers, yet under the complex and congested construction environment, GPS sensors sometimes are not reliable due to weak or lost satellite signals (especially when the construction sites are located in a dense urban area). Secondly, place fiduciary makers and applying quick response (QR) codes can interpret location information of mobile devices. This type of marker-based method does not depend on any network/infrastructure pre-installation or any mobile sensors. It is free of mobile data noise/accumulated-error-draft and works well in both indoor and outdoor environments, yet the field workers have to attach visual markers on each interested construction object. Tagging hundreds or even thousands of objects with 2D visual markers is quite a time-consuming procedure, which renders this application often impractical on complex construction sites. The Model-based method suffers from accumulated drift errors, requires pre-reconstruction of the 3D point cloud model of the physical world, and does not typically scale well [3].

70.3 Related Work

The fast innovation and development of computer vision techniques over the past decade have led to new research on the application of image-based localization methods for marker-less mobile systems. Some researchers have focused on visual SLAM (Simultaneous Localization and Mapping) for mobile systems by using parallel threads of tracking and mapping.

SLAM was originally proposed by Durrant-Whyte and Leonard [4] based on the earlier work of Smith et al. [5]. It solves the problem of generating and updating 3D maps of an indoor/outdoor environment while simultaneously tracking the location of the users within the environment. SLAM is a localization approach that tracks the spatial position and orientation of the agent with respect to its surroundings rather than one particular algorithm [6]. While doing location tracking, it constructs 3D maps (point cloud) through triangulating the detected features from the images of the environment.

Most modern simultaneous localization and mapping systems are based on tracking a set of visual features detected in image frames acquired from video cameras. Then, those tracked feature points can be utilized to triangulate the spatial position and at the same time, the estimated point locations can also be used to compute the camera orientations (spatial pose) within that observed environment. Even with a single camera sensor, by combining the measurements and tracking different feature points over multiple keyframes, it still has the capability to recover camera pose and position with high accuracy.

In this paper, we implemented a vision-based localization solution that can localize and track the position and orientation of handheld users on construction job sites in a near real-time manner on mobile devices. The approach does not call for any pre-installed fiduciary markers. It mainly includes 3D point cloud map generation, image-based localization, and alignment of point cloud with BIM model procedures. All these computing procedures are implemented and running on a remote server. Initially, a video will be acquired using the built-in camera of a mobile device by scanning around the job site by an on-site worker. Next, 3D point cloud map of the job site is to be generated from the acquired video stream by executing the mapping process. At the same time, localization process will also be executed by tracking the query image frame received from the end user. Afterward, the server will send the location/orientation results against the 3D map back to the end user. Finally, users can also align the 3D point cloud maps with their BIM models remotely on the user client for retrieving global location information as well as facilitating the identification, processing, and communication of discrepancies between actual and expected construction performances for project managers.

70.4 Problem Statement and Objective

Efficiently and accurately localizing and tracking position/orientation of onsite personnel will reduce both working time and efforts for providing on-demand access to project information. SLAM algorithm has significant potential to achieve onsite localization and tracking goals with high performance and fewer prerequisites (compared with GPS, fiduciary marker-based, model-based methods). Due to the availability of inexpensive, high-resolution and "point-and-shoot" cameras on mobile devices, it provides a rich source of information pertaining to construction facilities (images, videos, and aerial photos etc.) and makes it possible to apply SLAM on construction sites. This paper aims to identify the feasibility as well as test the performance and efficiency of applying the SLAM-based method to the task of tracking and localizing handheld users for construction field reporting on mobile devices.

70.5 Implementation and System Description

70.5.1 ORBSLAM-Based Localization System

In this research, an image-based localization prototype based on ORBSLAM algorithm [7–9] was developed with a client-server structure. It includes a mobile client designed on Android platform, an HTML-based (web viewer-based) client (this client has platform independence, it can be utilized on any device (e.g. IOS, Android, Windows phone and desktop etc.) that has a web browser) and a remote Linux server.

The principle of the localization system is illustrated in the flowchart as shown in Fig. 70.1. Firstly, end users can take photos/videos or select existing images/videos on the client and upload them to the server. Next, the server processes received image frames and generate a local 3D point cloud map by executing the mapping procedure. Afterward, the user can send query images (either real-time captured or stored on the mobile device) to the server. Then, the server localizes the query images against the 3D map. Moreover, users can align the 3D point cloud with BIM model by simply selecting at least 4 couples of corresponding points from the point cloud and BIM models on the client. Once alignment completes, the server sends the transformed global location information (aligned with BIM model) back to the client users.

70.5.2 Improved Mapping and Localization System

In order to apply the approach to facilitate the process of field reporting, the field reporters (end users) should not only be able to map the site on a specified time manner (daily, weekly, it actually depends on the construction progress), but also be able to localize themselves at any time, any location when doing field reporting. However, the SLAM method does the localization simultaneously with mapping by using successive image frames, yet it is not capable of doing post-localization. Therefore, the original SLAM algorithm does not fulfill the requirement of conducting field reporting on construction sites.

To overcome this drawback, we proposed an improved ORBSLAM-based mapping and localization system that has a client-server structure. This structure is designed to enable the field reporters (end users) to get quick (near real-time) response from the server whenever they send a query frame from the client.

The improved localization system includes a mobile client designed on Android platform, an HTML-based client and a remote Linux server, when the client connected with Wi-Fi or local server network, the users can do onsite video recording by using built-in camera of their mobile devices. Mapping process can be done once the video collection completes. The generated 3D point cloud maps then can be stored on the remote server. Once the users take new query images and send them to the server, the localization system will immediately get them localized and send the location information back to the users. Figure 70.2 shows the user client on Android platform of the improved localization system.

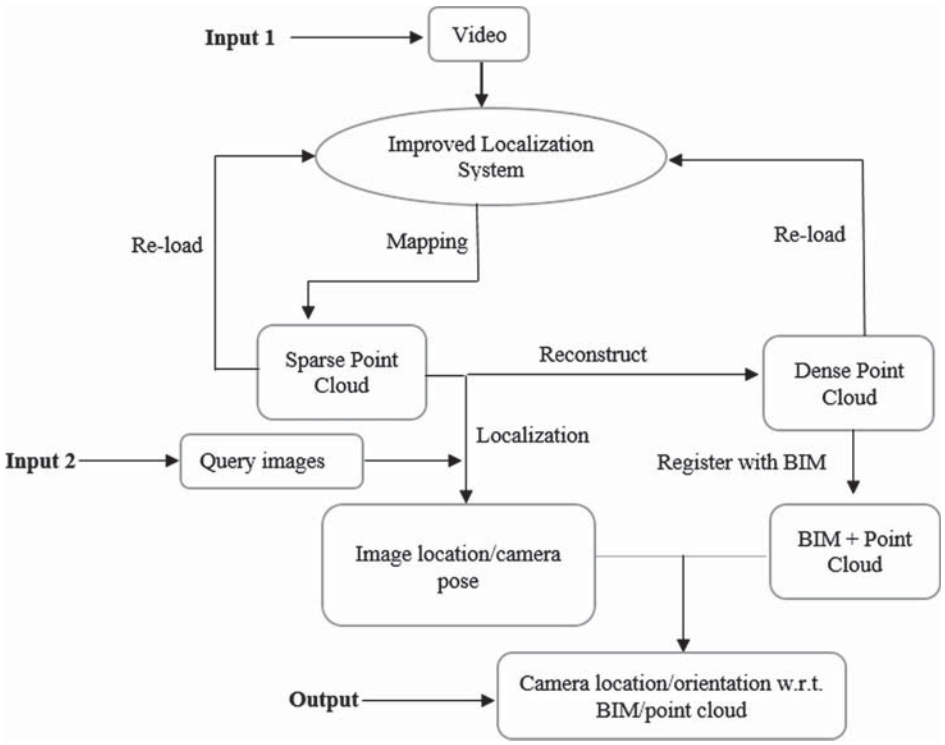

Fig. 70.1 Flowchart of mapping and localization prototype

70.6 Case Study

70.6.1 Data Collection

The data used in this study was collected in the Newmark Civil Engineering Lab (NCEL) at the University of Illinois at Urbana-Champaign. A 1 min 52 s video and ground truths were collected. The images in Fig. 70.3 show the scenes of NCEL.

70.6.2 Ground Truth

Ground truth data in this experiment was collected by the Google Project Tango tablet. It has motion tracking and area learning functions that enable users to track the device's position and orientation within a detailed 3D environment. The reason why we use Tango to generate the ground truth data is that the absolute ground truth for the NCEL dataset is unattainable. Furthermore, based on the performance evaluation of Project Tango conducted by Roberto et al. in 2016, for indoor scenarios, the Tango's precision of motion tracking is within 2 cm at a 95% confidence level [10]. According to their results, we can utilize the motion tracking data produced by the Tango as a reference to help evaluate the accuracy performance of our proposed localization solution.

In Fig. 70.4, the blue lines are the recorded 3D trajectories of the device. And the 3D point cloud data was generated by exporting points from the 3D mesh model reconstructed by the Project Tango.

70.6.3 Data Testing

To process the collected data, the recorded video was first input into the algorithm and processed on the server. After specified the settings, the algorithm processed the video data at a 10fps rate. As shown in Fig. 70.5, the green grids in the left picture are visual features that were real-time detected and tracked in each image frame. On the right side, the blue triangles

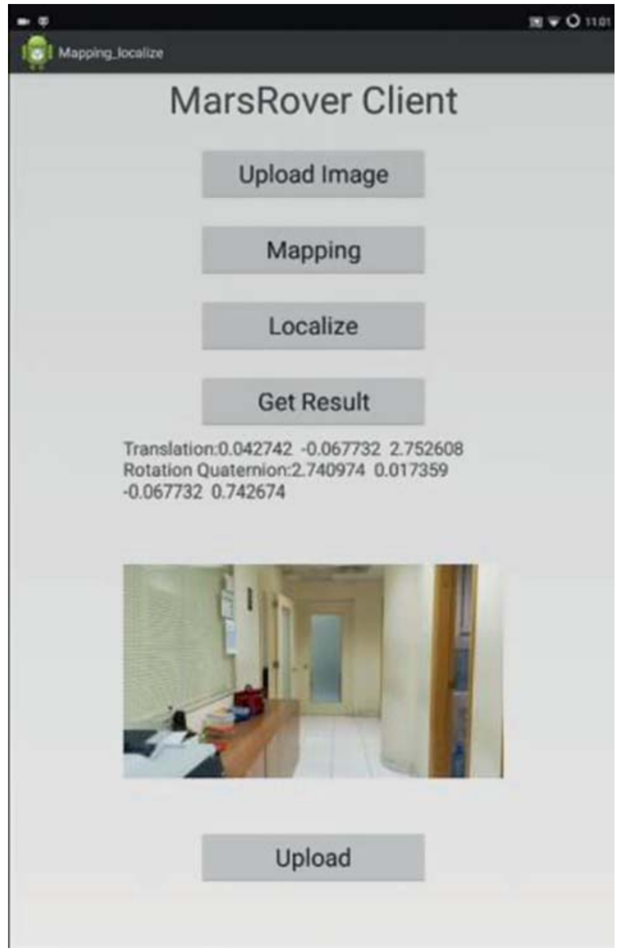

Fig. 70.2 Android client of the improved localization system

are the calculated poses of the device. Also, the generated 3D sparse point cloud is shown in the right window (red points of the point cloud were triangulated by using visual features within the current image frame, while the black points were generated without using the features in the current image frame).

Since the localization results are calculated against the 3D point cloud generated during mapping process and it is based on local coordinate system (not in real-world scale). Therefore, an improved localization system was developed, in the system, we implemented the interface for aligning 3D point cloud maps with BIM or 3D mesh models by manually selecting 4 couples of corresponding points on the web-based client. After the alignment of 3D point cloud maps and the BIM or 3D mesh models, the location information was transformed to the model-referred coordinate system and would be sent back to the end users.

70.6.4 Experimental Results and Accuracy Evaluation

After the SLAM process, a set of 30 image frames was down-sampled from the NCEL dataset to evaluate the accuracy of the localization system. Spatial location of the camera device can be represented by its 3D trajectory, in which 3D position and spatial orientation are represented by (Xi, Yi, Zi) coordinates and (q1, q2, q3, q4) quaternions respectively. In order to evaluate the accuracy of the localization results in the experiment, we need to compare the 3D position and orientation separately with the ground truths.

For evaluating 3D position accuracy, location errors are considered as the spatial distances of the corresponding points between the experimental results and ground truths. Three-dimensional distances can be calculated by $D_i = \sqrt{(Xi - Gxi)^2 + (Yi - Gyi)^2 + (Zi - Gzi)^2}$, where Xi, Yi, and Zi are the 3D spatial coordinates of the experimental results, while Gxi, Gyi, and Gzi are 3D spatial coordinates from the ground truth data.

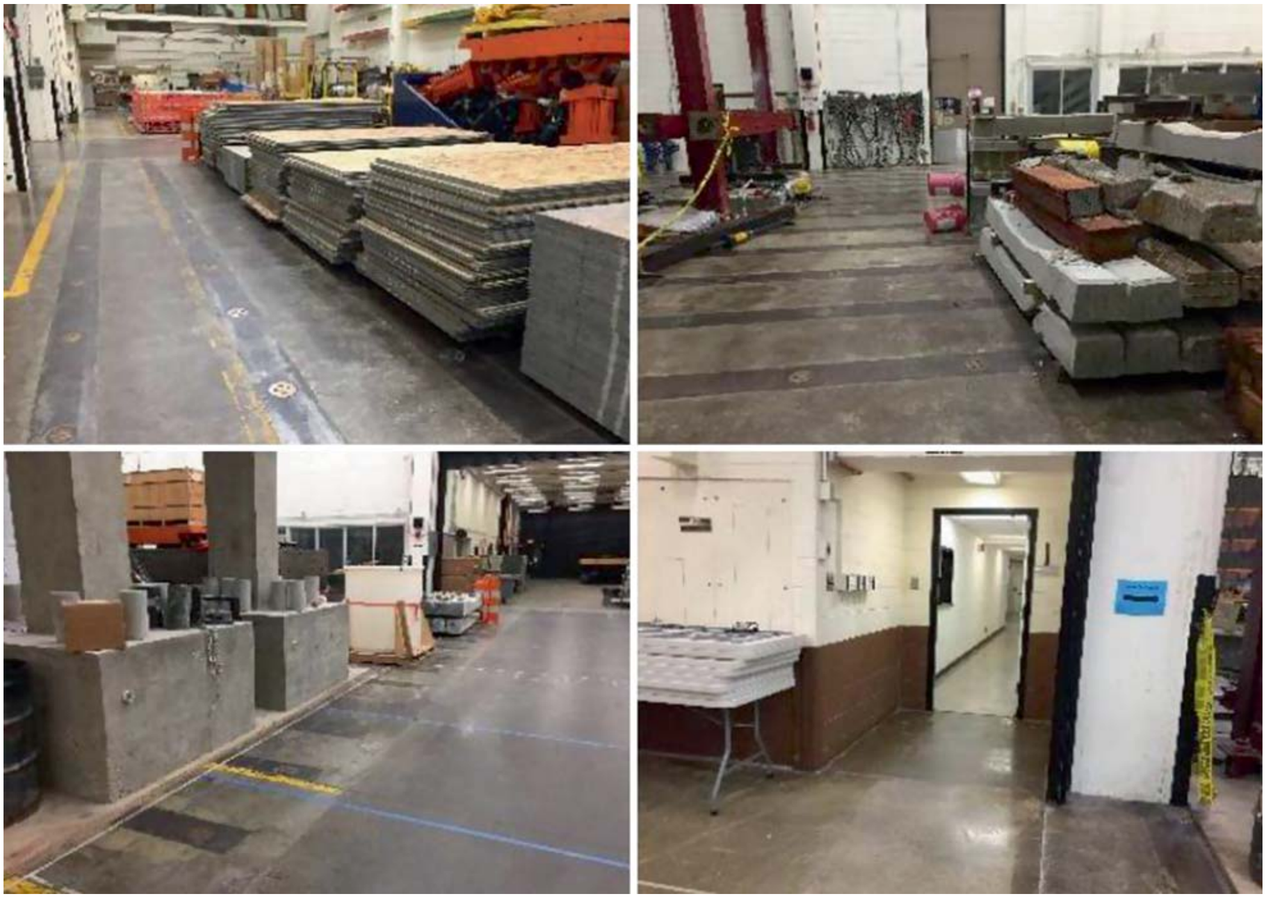

Fig. 70.3 Newmark civil engineering lab (NCEL) site

We calculated the Root Mean Square Error (RMSE) and standard deviation (STD) of the spatial distances for the experimental results by using the equations below.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} Di^2}{n}}$$

$$\mathrm{STD_s} = \sqrt{\frac{\sum_{i=1}^{n} (Di - \bar{D})^2}{n-1}}$$

(Here, n is the number of tested image frames, it is 30 in the experiment. STDs is the sample standard deviation).

To evaluate the accuracy of spatial orientation, we need to compare the quaternion matrices between the experimental results and the ground truths. The angle between two quaternions can be used to measure the differences in spatial orientation. When the angle ɵ between two quaternions is zero (in which case, there is cos(ɵ) = 1), which means the two quaternions are oriented in the same direction. Therefore, the Mean Absolute Percentage Error (MAPE) of the angle of corresponding quaternions between the experimental results and ground truths can be calculated by the equation below.

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \cos\left(\frac{\mathrm{Qti} \cdot Qgi}{|Q\mathrm{ti}| \cdot |Qgi|} \right) - 1 \right|$$

(Qti is the quaternion matrix from experimental results, and Qgi is the quaternion matrix from ground truths).

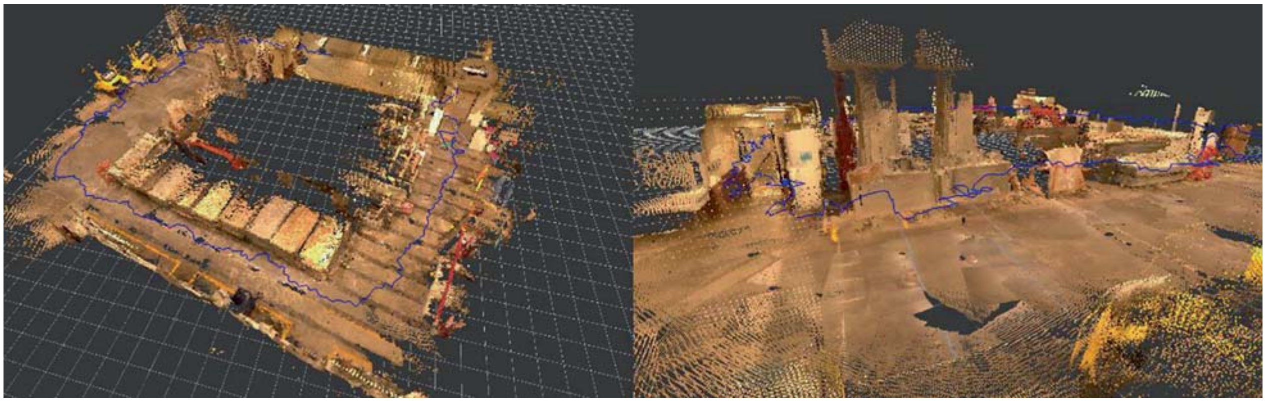

Fig. 70.4 Ground truth data of NCEL dataset

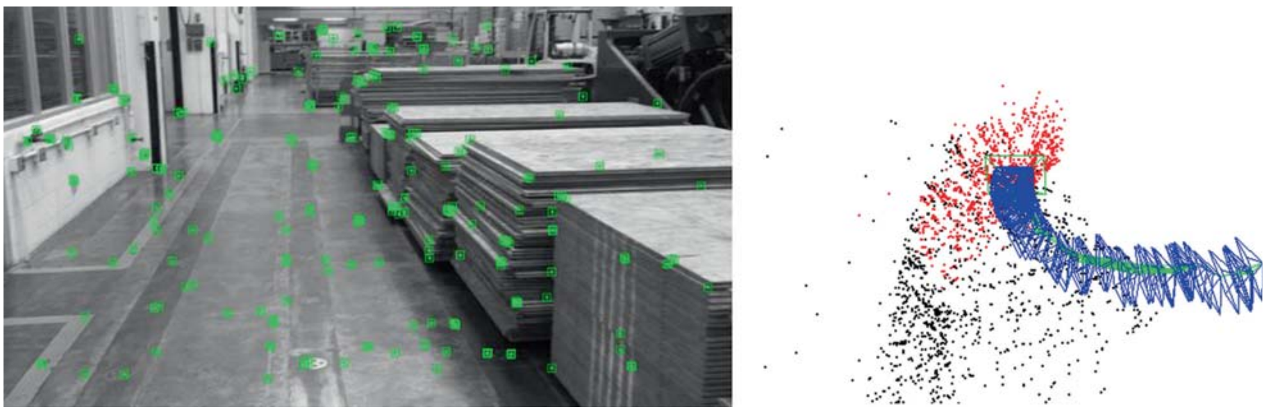

Fig. 70.5 Data testing of the localization system

In the experiment, the root mean square error of 3D position was calculated to be 0.03274 m, while the standard deviation is 0.00882 m for the testing dataset. For spatial orientation, the mean absolute percentage error is 0.001348, which is 0.1348% for the tested orientation results.

From the above calculation, we can see that the accuracy performance of the 3D position results is around 3.27 cm for the localization. According to the evaluation on Project Tango conducted by Roberto et al., we can claim that the proposed localization system can achieve a 3.27 ± 2 cm precision performance for indoor real-world scenarios. The spatial orientation results have a 0.1348% error, in which we used cos(ɵ) as the observations of spatial orientation to compare experimental data with ground truths. Through the accuracy evaluation in our laboratory, it reveals that the proposed localization solution in this study holds a great potential for construction onsite workers to help them get quick and highly accurate location information.

70.6.5 System Efficiency

The system localization running time was also recorded and shown in Fig. 70.6. The average processing time is around 1.071 s, which is in a near real-time manner.

70.7 Conclusion and Future Work

A mapping and localization solution with a client-server structure was proposed in this paper. It includes a mobile client designed as a smartphone application, a web-based client, and a remote computing server. A case study was conducted to evaluate the accuracy and efficiency performance of the proposed solution. According to the evaluation results, the solution

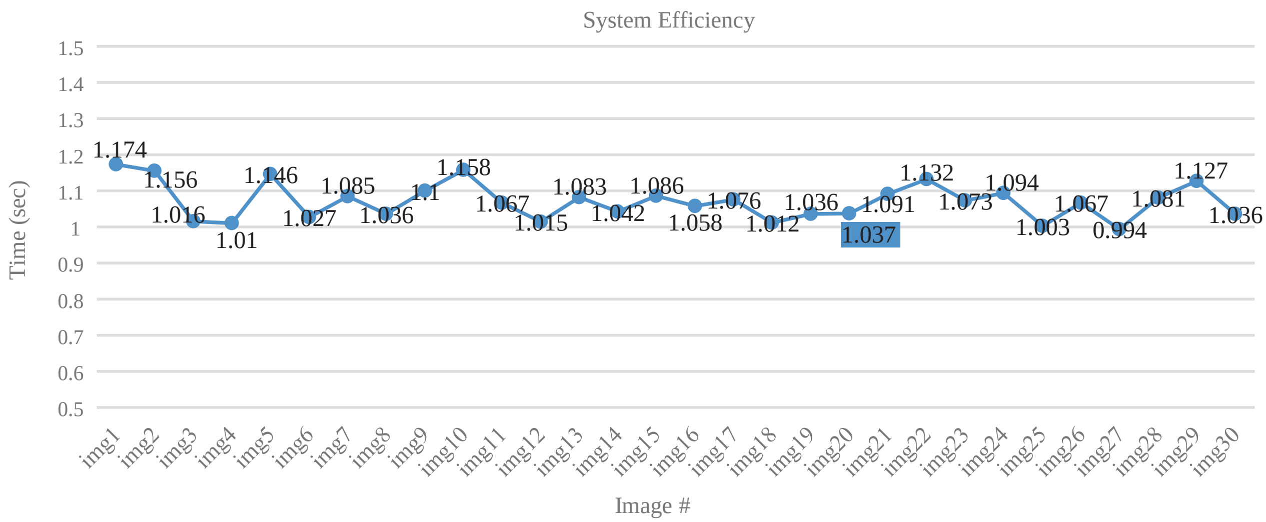

Fig. 70.6 System running time

proposed is potential to be applied to help construction field workers to obtain location information in near real-time. It will help to boost the efficiency of location-based tasks on construction sites, such as field reporting and progress monitoring.

The proposed solution still has some limitations that call for further research. Despite the fact that the system achieved near real-time localization, the mapping process is executed on the server side which means that the server has to wait till the users complete video recording and uploading processes that actually decreases the efficiency of the system. Also, approaches for updating 3D maps that are needed for tracking changes in dynamic construction environment were not discussed in this work. Furthermore, since the system is in a client-server structure, both the server and the user client need to connect to a wireless network (Wi-Fi, LTE, etc.). In future work, the authors will further evaluate the performance of the system and obtain feasible approaches to reduce the accumulated drift error produced during the localization procedures.

References

1. Golparvar-Fard, M., Peña-Mora, F., Savarese, S.: Application of D4AR—A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. ITcon **14**, 129–153 (2009)
2. Golparvar-Fard, M., Peña-Mora, F., Savarese, S.: Integrated sequential as-built and as-planned representation with D4AR tools in support of decision-making tasks in the AEC/FM industry. J. Constr. Eng. Manage. **137**, 1099–1116 (2011)
3. Bae, H., Golparvar-Fard, M., White, J.: Image-based localization and content authoring in structure-from-motion point cloud models for real-time field reporting applications. J. Comput. Civ. Eng. **29**(4), B4014008 (2014)
4. Leonard, J.J., Durrant-Whyte, H.F.: Mobile robot localization by tracking geometric beacons. IEEE Trans. Robot. Autom. **7**(3), 376–382 (1991)
5. Smith, R., Self, M., Cheeseman, P.: Estimating uncertain spatial relationships in robotics. Autonomous Robot Vehicles, pp. 167–193. Springer, New York (1990)
6. Thrun, S., Leonard, J.J.: Simultaneous localization and mapping. In: Springer Handbook of Robotics, pp. 871–889 (2008)
7. Mur-Artal, R., Tardós, J.D.: Fast relocalization and loop closing in keyframe-based SLAM. In: IEEE International Conference on Robotics and Automation (ICRA), May 2014, pp. 846–853 (2014)
8. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Rob. **31**(5), 1147–1163 (2015)
9. Mur-Artal, R., Tardós, J.D.: Orb-slam2: An open-source slam system for monocular, stereo, and RGB-D cameras. IEEE Trans. Rob. **33**(5), 1255–1262 (2017)
10. Roberto, R.A., Lima, J.P., Araujo, T., Teichrieb, V.: Evaluation of motion tracking and depth sensing accuracy of the tango tablet. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 231–234 (2016)