Full-body Pose Estimation for Excavators Based on Data Fusion of Multiple Onboard Sensors

Jingyuan Tang[a], Mingzhu Wang[b]*, Han Luo[ac], Peter Kok-Yiu Wong[a], Xiao Zhang[a], Weiwei Chen[d], Jack C.P. Cheng[a]*

a Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

b School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, United Kingdom.

c China Three Gorges Investment Management Co., Ltd, Shanghai, China

d Laing O'Rourke Centre of Construction Information and Technology Laboratory, Trinity College, University of Cambridge, Cambridge, United Kingdom.

*Corresponding Authors

Abstract.

To reduce machine-related accidents on sites, automatically monitoring the full-body poses of operating heavy machines is crucial. Conventional pose estimation systems relying on homogeneous sensors are vulnerable to negative environmental impacts, leading to inaccurate and unstable estimation of machine states. Hence, a full-body pose estimation framework is proposed for excavators, with a data fusion strategy to utilize different types of onboard sensors for enhanced accuracy and robustness. Specifically, a non-invasive onboard visual-inertial sensor system is designed for data fusion. Then, through competitive and complementary data fusion, the keypoints describing the full-body poses of the excavator are tracked in 3D space. Especially, an EKF-based localization algorithm is developed for optimized multi-keypoint tracking, which is verified to improve the accuracy and robustness of pose estimation by a real-world excavator case study. The proposed sensor-fusion method can effectively improve operational safety, by accurately monitoring the motion of heavy machines operating on construction sites.

Keywords. Data Fusion, Visual-inertial Sensor System, Pose Estimation, Construction Safety, Excavator Operation, Construction Machine

1. Introduction

The construction industry has been regarded as one of the most dangerous industries. According to the Occupational Safety and Health Statistics Bulletin published in 2021 by the Labour Department of Hong Kong [1], the construction industry had the highest accident rate and numbers of fatalities among all industry sectors in the past decade. In China, 904 workers died in construction safety accidents in 2019, up 7.26% year-on-year [2], with an average of 2.5 fatalities per day. In addition to casualties, these construction accidents have resulted in significant financial loss for employers, including medical costs, worker's compensation expenses, losses of project delay, etc. [3]. Therefore, it is important to address the construction safety issues and prevent potential dangers on construction sites.

In particular, operation of heavy construction machines constitutes a major cause of occupational hazards on construction sites. In 2020, contacting with or being struck by moving machines was reported as the second most common source of construction accidents in Hong Kong [4]. Occupational Safety and Health Administration (OSHA) [5] in the U.S. has also suggested struck-by machines as one of the top four construction hazards causing over 60% of construction-related deaths. In addition to directly causing casualties, the unsafe operations of a construction machine may also damage buried underground pipelines, and endanger other public and private facilities, pedestrians, and nearby residents. In order to avoid these accidents, in addition to training operators, external intervention measures are also needed. It is therefore necessary to monitor the operations of heavy machines on a construction site to prevent potential dangers, as well as improve operational safety and productivity. Traditionally, monitoring the operations of construction machines relied on inspector observing on site or watching a video captured by surveillance cameras [6], but such manual monitoring is labor-intensive and error prone work and is subjected to the inspector's reaction and experience. Therefore, automated solutions of construction machine monitoring are necessary to enable more precise and proactive operational safety management.

In the early stages, the automated operation monitoring of construction machines focuses on locating the machines on a two-dimensional (2D) map by localization technologies [7-9]. However, the vague information is not sufficient to adequately describe the working status of heavy machines on construction sites. It is observed that in construction activities the heavy machines (e.g., excavators) rarely change in locations, but their articulated parts, consisting of multiple movable independent components are operated in 3D space and form complex poses. Excavators are the most typical of such articulated equipment. An excavator has four movable components (i.e., a cabin, a boom, an arm, and a bucket) and, compared to other heavy machines such as trucks and bulldozers, an excavator has a higher degree of structural freedom, giving it a much greater range of motions and complexity in poses. Compared to varying locations, the changing pose of the excavator is more likely to make collisions with surrounding facilities, pedestrians, and vehicles to threaten operational safety. Hence, tracking the current 3D poses of articulated construction machines is essential and forms the basis of automated operational safety monitoring.

Recent studies have explored using only homogeneous sensors to track the motion of articulated construction machines. Both visual (e.g., cameras) [10, 11] and non-visual sensors (e.g., inertial measurement units (IMU)) [12] have been used to effectively estimate the (partial or full-body) poses of excavators. However, these pose estimation systems utilizing homogeneous sensors are unavoidably limited by environmental interferences and noises on construction sites, and consequently, cause inaccurate and unreliable descriptions of the pose, which is extremely dangerous for operational safety monitoring. To address the problem, the data from different sensors should be fused to improve the survivability of the pose estimation system under different conditions and optimize the description of excavator motions. Unfortunately, there is no effective full-body pose estimation approach for articulated construction machines by fusing data from multiple sensors.

This study therefore proposes employing data fusion a full-body pose estimation framework for monitoring machine in 3D. In the framework, first of all, a non-invasive onboard multi-sensor system comprising a stereo vision module and IMU sensors mounted on the machine — in our study, an excavator — is developed to track the machine's motion and collect data regarding its poses. With the various onboard sensors now in place, data can be fused competitively and complementarily, and through this data fusion, multiple keypoints on the body of the machine can be tracked by a developed multi-keypoint localization algorithm based on Extended Kalman filter (EKF), and then be combined to form a full-body 3D visual of the position and pose to have enhanced accuracy and robustness. The proposed approach provides the theoretical basis for developing an accurate and robust 3D full-body pose estimation of excavators on real construction sites to monitor the motions of machinery and improve operational safety.

The rest of the paper is organized as follows: Section 2 reviews relevant research on methods of tracking the movements of construction machines. Section 3 describes the data fusion–based full-body pose estimation approach proposed by this study. Section 4 illustrates tests that validate the approach, and Section 5 concludes with the research's contributions and limitations.

2. Related Works

This section reviews and evaluates relevant research on the pose estimation methods for construction machines using both homogenous and heterogeneous (multiple) sensors.

2.1. Pose Estimation of Construction Machines Based on Homogenous Sensors

Pose estimation refers to describing the spatial orientation and motion of (construction) machines. In previous studies, using homogeneous sensors, including visual and non-visual sensors, is common when tracking the motion states of machines.

Visual sensors such as digital cameras deployed near the machine and surveillance cameras mounted on site, capture the images with geometry and color information to record the motions of construction machines. Marker-based pose estimation attaches fiducial markers to the machine component to be estimated. An optical camera is used to monitor the fiducial markers, and to estimate their orientations which represent the motion states of the estimated component [11, 13, 14]. Although relying on markers, these methods help to develop a low-cost, high-deployment efficiency, and fast-recognition pose estimation system for construction machines. Additionally, other studies focus on using unmarked image processing to remove the limitation of marker recognition when tracking the motions of construction machines. For example, Soltani et al. [15] tracked the partial motions of an excavator by extracting the 2D skeleton. Multiple vision-based excavator parts' detectors, which were trained at different angles through synthetic images, were used to estimate the partial pose of the excavator by the skeletonization of each component in the foreground. Furthermore, to reduce the workload of training multiple detectors and improve the accuracy, Luo et al. [10] developed an end-to-end deep learning approach to estimate the full-body poses of excavators. The images collected by a surveillance camera are labelled with pre-defined keypoints of the machine, based on which three architectures of deep learning networks are trained to estimate the full-body pose of an excavator. In addition to monocular cameras, the stereo visual module can also be used in the pose estimation of construction machines. Soltani et al. [16] presented a stereo vision system with a long baseline on a large construction site to estimate the motions of excavators. The 3D pose of the machine was computed with 2D skeletons of partial excavator from each camera which is involved in the stereo vision system.

Using visual sensors and computer vision technology can effectively develop a low-cost and user-friendly pose estimation system, but it still has obvious disadvantages: Besides the instabilities caused by insufficient illumination and limited field of view, there are always obstructions of views on dynamic and complex construction sites which affect the accuracy of vision-based pose estimation [6]. Specifically, the moving machines and workers usually block the monitoring object (e.g., fiducial markers or joints), and render the pose estimation system lose its tracking target.

In addition to visual sensors, non-visual sensors have also been utilized to estimate the poses of construction machines non-invasively. Precision measuring equipment (e.g., LiDAR [17, 18]) and high-precision localization technologies (e.g., ultra-wideband (UWB) real-time location system (RTLS) [19]) can provide the location information of the keypoints to be estimated on the machine, which directly describe its motions. Although great accuracy can be achieved using these devices, the high price of these devices makes them inoperable in the construction industry. Current research has made attempts to use low-cost inertial measurement units (IMU) to estimate the poses of construction machines. IMU sensors can be installed on the surface of a movable component to record its rotation states in space [20-23]. Through kinematics modeling of construction machines, the rotations of different components can be integrated to describe the full-body pose of the machine [12]. The study on IMU-based pose estimation method claimed that using IMUs can effectively provide a spatial description of the full-body pose of a construction machine with an accuracy of 90%.

However, for non-visual sensors, the unmodeled noises and deviations are unavoidable due to the intrinsic characteristics of sensors and the negative influences from the external environment, which lead to inaccurate and unstable machines pose estimation in practical applications [24]. For example, in the IMU-based pose estimation, when the temperature rises during operation, the performance of the IMUs decreases, which causes systematical problems including data loss and uncontrollable measurement errors.

Overall, although both visual and non-visual sensors can be used to describe the poses of construction machines, due to the limitations and errors that are unavoidable for any type of measurement, using only homogeneous sensors in pose estimation is instable and inaccurate in practice. Especially, for operational safety monitoring, any deviates that render the monitoring system abnormal or fails to work is dangerous. Therefore, it is necessary to use a multi-sensor (heterogeneous) system to make the information obtained from different sensors (i.e., visual or non-visual sensors) complement or compete with each other, so as to ensure the stability of the full-body pose estimation and improve its accuracy.

2.2. Pose Estimation of Construction Machines Based on Heterogeneous Sensors

Using a heterogeneous sensor system for pose estimation requires fusing data from different sensors. Data fusion can be done complementarily or competitively [25].

For complementary fusion, the mutually-exclusive data from different sources are integrated to extend the spatial and temporal coverage of the sensors, then appended to each other to piece together a full picture. Currently, using multi-sensor in pose estimation of construction machines has had only a smattering of studies, and mainly focuses on fusing data complementarily to get abundant pose-related information. Kim et al. [26] for example present a multi-sensory system to track the position in 3D of the cutting edge on a bulldozer's blade. This system complementally fuses orientation and 2D location provided by motion sensors and RTK GPS to estimate the spatial motion status of the end effector with errors no more than 30 mm. Additionally, In Soltani et al. [16]'s stereo-vision-based pose estimation system, they fused locations from GPS and images from cameras complementarily to decrease processing efforts of excavator detection and improve the accuracy. However, relying on the integration of mutually-exclusive data cannot reduce the uncertainty of the pose estimation system, so the shortcomings of using homogenous sensors mentioned in Section 2.1 cannot be overcome. Hence, to improve the accuracy and robustness of the pose estimation for construction machines, in addition to complementary fusion, competitive data fusion is also needed to be used in the pose estimation of construction machines.

In competitive fusion, an object's motion (e.g., movements of the boom and the arm of an excavator) is tracked redundantly (i.e., the same component/part tracked by more than one sensor), and the description of the object, in the end, is optimized by the competitive data. Especially for operational safety monitoring, due to the requirement of locating potential hazards, the poses of construction machines should be directly optimized at the level of 3D locations of pre-defined keypoints for accurate and reliable representations of motions. In the manufacturing industry, competitive fusion with a multi-sensor system has given excellent performances in tracking a single point of a manipulator. According to the dynamics model proposed by Moberg et al. [27], Axelsson et al. [28] present an EKF-based method to estimate the tool position of a robot with two degrees of freedom. The accelerations of the robot tool and dynamics parameters (i.e., motor torques and motor angles), which are from different sources, are fused in their proposed method. However, considering the ease with which the measurement devices need to obtain the required parameters non-invasively without making extensive modifications [29], the data fusion method based on dynamics model cannot satisfy the needs of applications for construction machines, because the parameters required by dynamics models are difficult to obtain using non-invasive sensors. Specifically, many off-the-shelf machines in practical require pose estimation system which can be directly mounted on surfaces without any modification inside the machine, as it can avoid refurbishing outdated machines, reducing both labor and financial costs for users. Therefore, the non-invasive sensor-fusion technologies based on a kinematics model should be the practical exploratory direction of the operational safety monitoring for construction machines. Liu et al. [30] uses a Kalman filter (KF) and multi-sensor optimal information fusion algorithm (MOIFA) to fuse the data collected by a multi-sensor system, which included a visual sensor and an angle sensor, and managed to improve accuracy by 38% ~ 78%. Ubezio et al. [31] conducts end-effector tracking on a nonlinear manipulator using sensor fusion techniques and a particular visual-inertial sensor suite. It proves to be more accurate and robust than homogenous sensor measurement on a complex machine. These previous studies show the ability of competitive data fusion with heterogeneous sensors to improve accuracy and reduce the uncertainty for single point (i.e., the end-effector) localization of a manipulator. However, for the articulated construction machine with multiple components (e.g., excavators), when monitoring its operational safety, the locations of multiple keypoints on independent components should be tracked simultaneously to comprehensively represent its pose. However, there is a lack of method on competitively fusing data from heterogeneous (multiple) non-invasive sensors to locate multiple pre-defined spatial keypoints on different movable components of a construction machine.

2.3. Research Gaps

According to the research reviewed in Sections 2.1 and 2.2, the research gap in existing pose estimation methods of excavators can be summarized as the following points:

- Instability and inaccuracy of existing full-body pose estimation based on homogeneous sensors for excavators.
- Lack of an accurate and robust multiple keypoints localization algorithm for excavators by fusing data from multiple sensors competitively.

It is therefore necessary to develop a full-body pose estimation framework for excavators based on a fusion of data collected from multiple onboard sensors, including competitive and complementary fusion. In this framework, a multi-keypoint localization algorithm should be designed for excavators to competitively incorporate data and provide pose information accurately and stably.

3. Methodology

As described in Section 2, introducing multi-sensor fusion into the pose estimation method is expected to improve the accuracy and robustness of motion tracking of construction machines. This study therefore proposes a full-body pose estimation framework based on data fusion of multiple on-board sensors for excavators, which is illustrated in Fig. 1. This proposed framework consists of two steps: (1) non-invasive on-board multi-sensor system and (2) full-body pose estimation of excavators based on data fusion. More details of the proposed study are given in the following sub-sections.

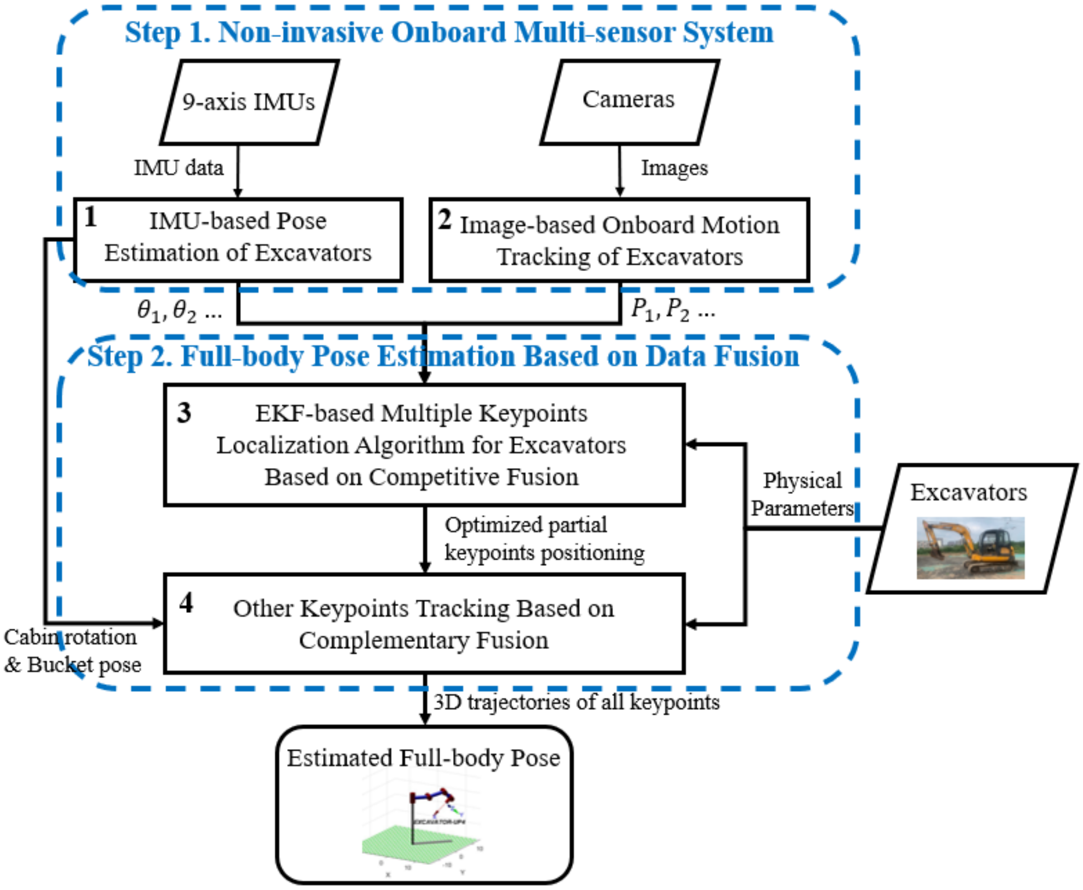


Fig. 1 Full-body pose estimation framework based on data fusion of multiple on-board sensors for excavators

The keypoints of an excavator are defined as the positions where the collision may occur in practice, including the end of each movable component and the rear edge, as well as the important connection point for transmitting motions. Fig. 2 shows the pre-defined keypoints of an excavator: K1 denotes the end of its cabin; K2 denotes the joint between the boom and the cabin, called the boom joint; K3 denotes the joint point between the boom and the arm — the arm joint; K4 denotes the joint point between the arm and the bucket, known as the bucket joint; and K5 denotes the end point of the bucket. These definitions will be used throughout the paper when the keypoints or pre-defined keypoints are mentioned without further elaboration. K1, K2, K3, K4, and K5 are coplanar.

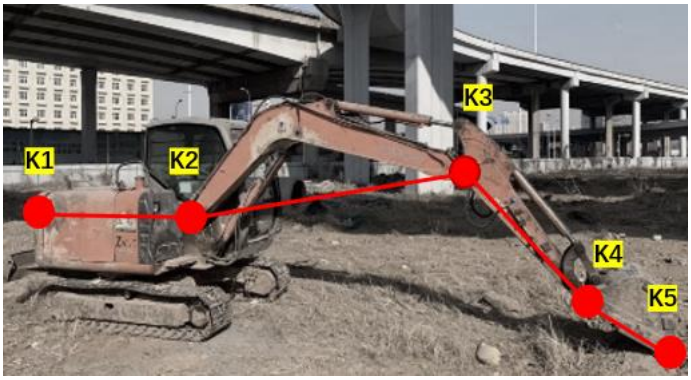


Fig. 2 Defining keypoints on an excavator

Five major reference frames are used in this proposed framework. They are:

(1) the sensor frame (x_b, y_b, z_b), which is attached to the IMU on the movable component of the excavator;

(2) the pixel frame (u, v), which is attached to the image, with the *u*-axis pointing to the right in the image's plane, the *v*-axis pointing down, and the origin located at the left corner of the image;

(3) the camera frame (x_c, y_c, z_c), which is attached to the camera with the *z*-axis pointing to the optical axis; the *x*-axis pointing to the right direction on the image plane; the *y*-axis pointing to the down direction on the image plane, and the origin located at the optical center of the camera;

(4) the projected 2D frame (x, y), which is attached to the camera with the *x*-axis pointing to the optical axis, the *y*-axis pointing up, and the origin being the optical center of the stereo vision module; and

(5) the world frame (x_w, y_w, z_w), which facilitates users to conduct further pose-related analyses and is determined based on the users' needs.

3.1. Excavator Pose Information Collection and Processing Based on A Developed Non-invasive Onboard Multi-Sensor System

In this step, a non-invasive on-board multi-sensor system is developed to collect pose information from two different data sources (i.e., IMUs and cameras) and to fuse the data. IMUs are attached to movable components of an excavator to estimate its poses. Simultaneously, a stereo vision module is installed on the cabin to track the trajectories of excavator keypoints based on a developed image-based onboard motion tracking method.

3.1.1. Sensor Selection

As discussed in Section 2.1, previous studies have demonstrated the characteristics and applicable scenarios of different techniques for estimating the poses of construction machines. Among those techniques, two types of sensors are predominant ones. One is inertial measurement unit (IMU), which has been widely studied and used in sensor fusion applications, because of its low cost, user-friendliness, quick response and not being susceptible to occlusion and illumination [24]. Another type is cameras as they can provide visual information directly without drift, based on which the position of the excavator's keypoints can be obtained with computer vision methods [10, 15]. Considering the above-mentioned complementary properties of IMUs and cameras, the proposed framework focuses on fusing data from both sensors, i.e., a visual-inertial sensor suit, where the angular data from IMUs and the visual information from cameras can complement each other to enable more accurate motion tracking.

3.1.2. IMU-based Pose Estimation of Excavators

As illustrated in Section 2.1, IMU sensors are installed on an excavator to collect angular data. The objective of this section is to obtain four types of information on angular sequences: (1) change of the joint angles between the cabin and the boom; (2) change of the joint angle between the arm and the boom; (3) the joint angle between the bucket and the arm, and (4) the cabin's angle of rotation. Such IMU-based pose information is obtained based on an existing method, developed by Tang et al. [12], the workflow of which is shown in Fig. 3.

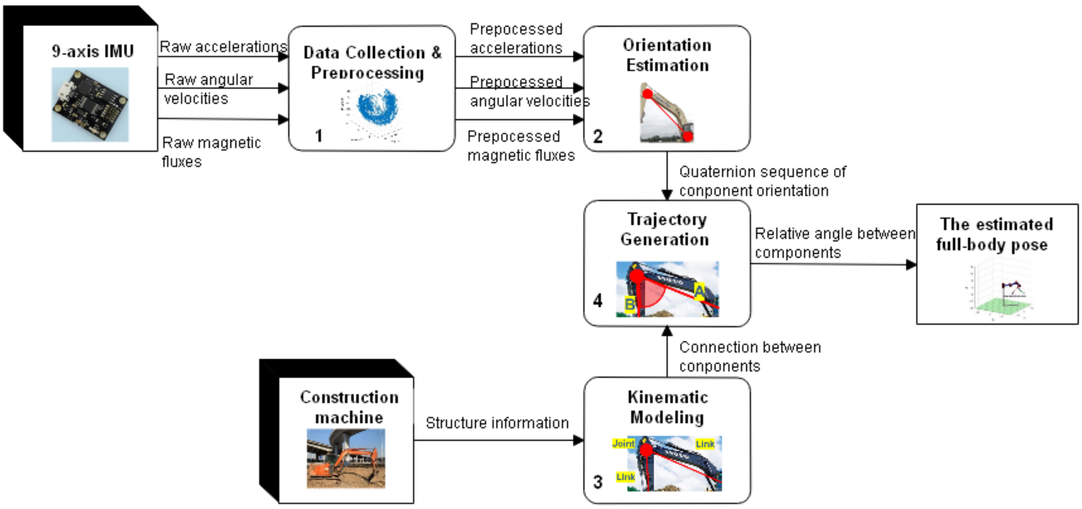


Fig. 3 Flow of information in the IMU-based Full-body Pose Estimation for construction machines[12]

9-axis-IMUs are attached to the surface of every movable component for the target excavator (i.e., cabin, boom, arm, and bucket), in order to collect three types of inertial data: (1) acceleration captured from the sensor's accelerometer, (2) angular velocity obtained from the gyroscope, and (3) magnetic flux collected from its magnetometer. First, raw data collected by the IMUs is preprocessed to remove noise caused by vibrations and other uncertainties intrinsic to the IMUs. Then, the de-noised IMU data is transformed to the orientation of each component for the excavator based on a quaternion-based drift-free orientation filter (e.g., Madgwick filter [32]). Afterwards, to specify the connection between the estimated orientation of each independent component as mathematical relationships, a kinematics model is built based on the structural information of the excavator. Finally, combining the estimated orientations of components and the kinematics model, the angular trajectories (i.e., the cabin's rotational angle and the relative angles of adjacent components) which can directly describe the pose of the excavator are generated using a developed quaternion-based method. The outputs of the angular sequences on the change of the joint angle between the cabin and the boom and the change of the joint angle between the arm and the boom need further processes. To obtain these sequences of angular changes, the relative angle of adjacent components at the time k is subtracted by that at time k–1. Consequently, four types of angular sequences required by the data fusion are obtained without drifts and partially modellable noises.

3.1.3. Image-based Onboard Motion Tracking of Excavators

In addition to angular data obtained from the IMUs, cameras are used as another data source for data fusion to collect visual information and track keypoints' positions of the target excavator. As discussed in Section 2.1, the major problem of existing computer-vision-based motion tracking methods for excavators is the frequent mutual occlusions between the target machine and obstacles on construction sites. To address the foregoing problem, we design an independent onboard system configured for the excavator. An additional advantage of the developed system is that all the required keypoints can be located only by obtaining the position of a single feature point. Compared to previous studies [10, 11] where the pose of the excavator needs to be estimated by identifying multiple points distributed in different components, our method can improve the deployment efficiency and reduce the computational cost. As illustrated in Fig. 4, the proposed image-based onboard motion tracking method consists of four components: (1) hardware setup; (2) camera calibration; (3) single feature point tracking; and (4) image-based keypoint estimation. The sub-sections explain the method in detail.

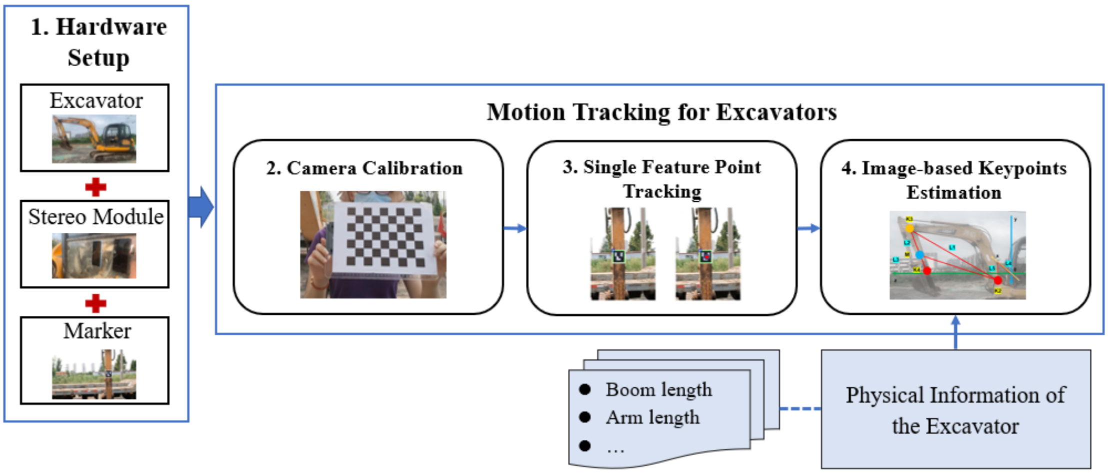
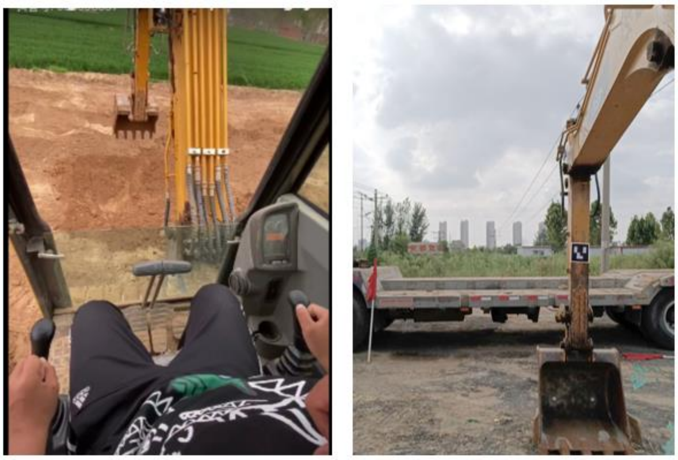


Fig. 4 Image-based onboard pose estimation method for excavators

Hardware Setup. The independent onboard method is designed with inspiration from the features of the operators' practical excavation works. Specifically, during digging and dumping, the operators pay more attention to the location of the excavator's arm, and they always ensure that the lower part of the arm can be seen without any occlusion, while the bucket is usually obscured by rock or soil. In addition, the operators intuitively estimate the current pose of the excavator using their eyes by observing the arm. According to such experience, it is found that if cameras are simulated as the operator's eyes and estimate the poses of the excavator like human, the problem of occlusions can be solved to a large extent. Hence, two cameras, which provide RGB and geometric information simultaneously, are used to build a stereo vision module in the proposed independent onboard method, which is mounted at the front of the cabin to simulate the operator's eyes. A marker is attached to the lower part of the arm to mimic the focus of the operator's eyes to facilitate estimating poses. The marker should be always in the view of cameras. Fig. 5 shows the actual operator's view and the view obtained by the stereo vision module, as well as the attached marker. As shown in this figure, due to the limited field of view, the cameras can only provide the positions of partial keypoints on the excavator, i.e., the boom joint (K2), the arm joint (K3), and the bucket joint (K4). As the bucket is usually blocked by soil and rocks during excavation, it is impossible to effectively provide the position of the end point of the bucket (K5). Since the cameras are installed on the cabin, it is also impossible to observe the position of the end of the cabin (K1). However, information on K1 and K5 will be obtained by methods introduced in Section 3.2.2.



(a) The actual operator's view. (b) The view of the stereo module.

Fig. 5 Comparison of (a) the actual operator's view with (b) the view as obtained by the stereo module

Camera Calibration. This is the process of obtaining intrinsic and external information about the camera and standardizing the image through estimating the camera's parameters. Camera calibration technologies have been quite mature, and this study adopts Zhang's method [33], which features a simple process with no professional-grade equipment, and is completed only by viewing a checkerboard with unknown orientations. When using a stereo vision module, in addition to calibrating each of the two cameras independently, the rotational and translational relationships between cameras also need to be established. This study uses the stereo calibration method by Hartley [34], which uses an essential matrix to show the relationship between the image pair normalized by the intrinsic and external parameters. After the camera calibration, images from the visual sensors are normalized and prepared for the feature point tracking.

Single Feature Point Tracking. Instead of requiring information of all keypoints, a single feature point is used to improve the efficiency of having to track multiple keypoints. The single feature point is defined as the centroid of the marker, and its coordinates are tracked in the camera reference frame based on the standardized images. First, the outer contour of the attached marker is detected. Although various types of markers are available in this method, in order to overcome the changing background on construction sites and enhance the stability of detection, binary square fiducial markers with their pre-defined libraries, such as ArUco [35], are selected in this study. After that, the feature point $(u_{centroid}, v_{centroid})$ in RGB images is calculated using moments, as shown in Eqs. (1) and (2).

$$u_{centroid} = \frac{\sum I(u,v)u_i}{\sum I(u,v)} \tag{1}$$

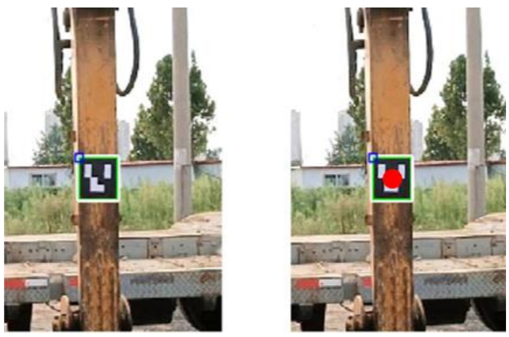
$$v_{centroid} = \frac{\sum I(u,v)u_i}{\sum I(u,v)} \tag{2}$$

where u_i and v_i denote the pixel coordinates of the *i*-th mass point along the *u* and *v* axes respectively; $I(u,v)$ denotes the density function related to the mass of each point in the contour. Fig. 6 shows an example of detected outer contour of a fiducial marker attached on the excavator and its centroid. Afterwards, the depth information of the feature point is extracted from the corresponding positions of the 3D map generated by the stereo vision module. Finally, combining the given pixel coordinates of the feature point and its depth information, the location of this point is projected to the camera reference frame based on basic camera model [36], using Eqs. (3) and (4).

$$x_c = z * (u_{centroid} - c_x)/f_x \tag{3}$$

$$y_c = z * (v_{centroid} - c_y)/f_y \tag{4}$$

where (x_c, y_c) shows the coordinates of the feature point in the camera reference frame in *x* and *y*-axes; *z* is the depth information of the feature point; c_x and c_y denote the optical center of the camera; f_x and f_y represent the focal length of the camera. All the camera parameters are included in the intrinsic matrix obtained in camera calibration.



(a) Detected outer contour of a fiducial marker (b) Centroid of the detected marker

Fig. 6 Detected outer contour of a fiducial marker and its centroid computed by moments.

Image-based Keypoints Estimation. Combining the location of the single feature point and physical parameters of the excavator, this step estimates, in the camera's reference frame, the coordinates of the target keypoints on the excavator (i.e., the boom joint (K2), the arm joint (K3), and the bucket joint (K4)). The major challenge of the developed method is that due to the large motion amplitudes of operating excavators and limited field of view of the camera, it is impossible to always keep each target keypoint in the field of vision of the cameras. In the proposed method, to ensure the marker attached to the lower part of the arm is always visible, the position of the arm joint (K3) cannot be directly observed in images. The location of the keypoint beyond the visual range (e.g., in a blind spot) is estimated through known information. A geometric decoder of excavators is designed to estimate the coordinates of the blind spot based on a given feature point and physical information of the excavator, and then further determine locations of all target keypoints in the camera reference frame. Fig. 8 shows details of the developed algorithm of the geometric decoder. First, six physical parameters of the excavator are manually measured in advance: (1) L1 — length of the boom, (2) L2 — distance from the joint point between the boom and the arm to the centroid of the marker, (3) L3 — length of the arm, (4) L4 — horizontal distance from the center of the camera to the boom joint, (5) L5 — vertical distance from the center of the camera to the boom joint and (6) L6 — depth from the center of the camera to the boom joint. Fig. 7 illustrates the physical information of an excavator, where the yellow point represents the blind spot, while the blue point M represents the centroid of the detected marker. Afterwards, K3 is estimated based on the structural relationship of the excavator. Eqs. (5) and (6) elaborate the basic principles of the blind spot estimation:

$$\overrightarrow{K2K3} = [0, -sin\angle K2 * z_{\overrightarrow{K2M}} + cos\angle K2 * y_{\overrightarrow{K2M}}, cos\angle K2 * z_{\overrightarrow{K2M}} + sin\angle K2 * y_{\overrightarrow{K2M}}] \quad (5)$$

$$K3 = [x_{K2}, normalize\left(y_{\overrightarrow{K2K3}}\right) * L1 + y_{K2}, normalize\left(z_{\overrightarrow{K2K3}}\right) * L1] \quad (6)$$

where$(x_{\overrightarrow{K2M}}, y_{\overrightarrow{K2M}}, z_{\overrightarrow{K2M}})$ shows the vector K2M; $normalize(x_{\overrightarrow{K2K3}}, y_{\overrightarrow{K2K3}}, z_{\overrightarrow{K2K3}})$ denotes the normalized vector K2K3; (x_{K2}, y_{K2}, z_{K2}) denotes the coordinates of the K2 all in the camera reference frame. Finally, according to the estimated blind spot K3 and the length of the arm (L2), the coordinates of the K4 can be computed. When locations of all target keypoints are recorded at each moment, the trajectories of these keypoints in the camera reference frame are drawn to describe the motion of partial excavator.

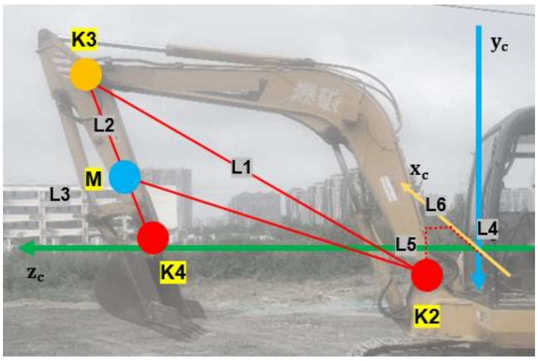


Fig. 7 Physical parameters of an excavator.

Algorithm 1 Geometric Decoder for Excavators

Input: Pt_K2(x_{K2},y_{K2},z_{K2}), Pt_M(x_M,y_M,z_M), K2K3(L1), K3K4(L3), K3M(L2),Video v (v_i is the i^{th} frame)

Output: Trajectory_K3,Trajectory_K4

Let K2 is the boom joint and its coordinates should be pre-defined as (L4,L5,L6). K3 is the arm joint and the blind spot need to be estimated in the algorithm. K4 is the bucket joint. M is the centroid of the marker attached to the arm.

i← 1 (Initialize the current frame ID)

total-Frame← GetVideoFrameNum(v) (Get the total frame number of v)

Trajectory_K3← 0 (Initialize the saved K3)

Trajectory_K4← 0 (Initialize the saved K4)

while i≤ total-Frame **do**

  (Estimate the coordinate of the blind spot(K3))

  K2M = Norm(Pt_K2-Pt_M) (Distance between the boom joint and the centroid of the marker)

  cosK2 = $(K2K3^2+K2M^2\text{-}K3M^2)/2\cdot$K2K3$\cdot$K2M

  sinK2 = sin(acos(cosK2))

  vector$_{K2K3}$ = $normalize$[$cosK2$(z_M-z_{K2})+sinK2(y_M-y_{K2}),-sinK2(z_M-z_{K2})+cosK2(y_M-y_{K2})]

  BlindSpot(K3)=[x_{K2},vector$_{K2K3}(y)\cdot$K2K3+y$_{K2}$, vector$_{K2K3}(x)\cdot$K2K3]

  Trajectory_K3.Add(K3) (Save the current K3)

  (Calculate the coordinates of K4)

  vector$_{MK3}$ = $normalize$[z$_{K3}$-z$_{K3}$,y_M-y_M]

  Pt$_K$4 = [x$_{K2}$, vector$_{MK3_{(y)}}\cdot$K3K4+y$_{K3}$,vector$_{MK3_{(x)}}\cdot$K3K4+z$_{K3}$]

  Trajectory_K4.Add(K4) (Save the current K4)

  i≤ i+1

return Trajectory_K3,Trajectory_K4

Fig. 8 Algorithm of geometric decoder of excavators to track target keypoints in the camera reference frame.

3.2. Full-body Pose Estimation Based on Data Fusion

As discussed in Section 3.1, data collected by IMUs and cameras in the onboard visual-inertial sensor system is used separately to measure the excavator's motion. However, two problems need to be further investigated. Firstly, measurements from different sensors are imprecise and unstable. Due to the negative influence of the sensors' intrinsic mechanical structure and the external environment, the unmodeled deviation and feature missing (e.g., stochastic noise and data loss) affect the accuracy of the measurements inevitably and increase uncertainty of the motion tracking system. Secondly, homogeneous sensors have limited spatial coverage, so they cannot provide thorough information on the full pose of an excavator. Specifically, for cameras, due to their limited field of view, it is difficult to measure the movements of the bucket and the cabin. To address such problems, this study proposes to fuse the IMU and camera data by a visual-inertial system. First, to improve the accuracy and robustness of the system, a competitive fusion is achieved in the articulated part of the excavator (i.e., boom and arm). A multiple keypoints localization algorithm is developed to combine the IMU and camera measurements competitively and find optimal estimations of the locations of the keypoints in the camera reference frame. After that, an effective complementary fusion is conducted with data at the cabin and the bucket to extend the spatial coverage of independent sensors and provide full-body pose information of the excavator.

3.2.1. The Developed Multiple Keypoints Localization Algorithm for Excavators Based on Competitive Fusion

An EKF (Extended Kalman Filter), a classical approach for non-linear stochastic system [37], is utilized in this study for competitive data fusion. An EKF linearizes non-linear systems using first-order approximation, and gives optimal results via a process with long iterative tuning. The EKF compensates for the limitations of using IMUs and cameras separately in motion tracking, so that the sensor fusion system has better performance than using a single type of sensors. The general EKF functions [37] are given. Let

$$x_{k+1} = f(\widehat{x_k}, u_k, w_k), w_k \sim N(0, Q_k) \tag{7}$$

$$y_k = h(x_k, v_k), v_k \sim N(0, R_k) \tag{8}$$

where $f(\cdot)$ is the state transition unction; x_k denotes a state vector; u_k denotes a known control input; w_k denotes the process noise, and v_k denotes the measurement noise; y_k represents the measurement vector; $h(\cdot)$ is the observation function, all in time *k*. The process noise w_k and measurement noise v_k are assumed as zero-mean white Gaussian noise with covariance matrixes Q_k and R_k, respectively. The EKF takes the first-order part of the Taylor expansion at its reference point as the approximation of the linear model and obtains the linearized description of the nonlinear system at time *k*. The prediction equations of the linearized system are given in Eqs. (9) and (10):

$$x_k^- = A(\widehat{x_{k-1}}, u_k) \tag{9}$$

$$P_K^- = AP_{k-1}A^T + Q \tag{10}$$

where A is the transition matrix, which is the partial derivative of $f(\cdot)$ with respect to the x at $\widehat{x_k}$; P denotes the variance of the predicted state estimate. The measurement update functions are shown as Eqs. (11) and (12).

$$\widehat{x_k} = \widehat{x_k}^- + K_k(y_k - H(\widehat{x_k}^-) \tag{11}$$

$$P_k = (I - K_k H_k)P_k^- \tag{12}$$

where the Kalman gain K_k is given as Eq. (13):

$$K_k = \frac{P_k^- H_k^T}{H_k P_k^- H_k + R} \tag{13}$$

H_k is the Jacobian matrix, which is the partial derivative of $h(\cdot)$ with respect to x at the prior state estimation $\widehat{x_k}^-$.

The state transition functions, and observation functions are built based on the specific motion modes of the excavator. It is noted that according to the characteristics of excavator motions, the keypoints tracking problem in 3D space can be projected onto a 2D plane to reduce the complexity of the functions and improve computational efficiency. Since the stereo vision module is installed on the cabin, no matter how the components move, including the rotation of the cabin, all the target keypoints can be projected onto a fixed 2D plane in the camera frame. Fig. 9 illustrates the excavator model and the projected 2D coordinates system. In the projected 2D plane, the boom joint (K2) is a fixed point, which can be determined by the physical parameters of the excavator. The arm joint (K3) and the bucket joint (K4) are moving according to the movement of different components.

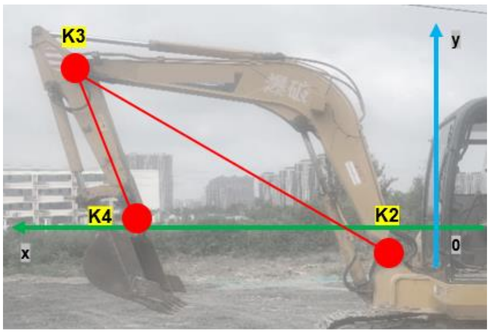


Fig. 9 Excavator model and projected 2D coordinates system.

In the algorithm, the locations of K3 and K4, which are directly observed by cameras, are inputted as measurements. The changes of relative angles estimated by IMUs (i.e., the change of the joint angle between the cabin and the boom, and the change of the joint angle between the arm and the boom) are used to predict the state estimations. The state vector is given as:

$$X_k = [x_3, y_3, x_4, y_4]^T \tag{14}$$

where (x_3, y_3) denotes the coordinates of K3, and (x_4, y_4) denotes the coordinates of K4, all in the projected 2D frame. According to the kinematic relationship of the excavator model, the state transition functions are given as follows:

$$x_3^k = x_2 + (x_3^{k-1} - x_2)cosu_1 - (y_3^{k-1} - y_2)sinu_1 \tag{15}$$

$$y_3^k = y_2 + (y_3^{k-1} - y_2)cosu_1 + (x_3^{k-1} - x_2)sinu_1 \tag{16}$$

$$x_4^k = x_3^k + (x_4^{k-1} - x_3^{k-1})cosu_2 - (y_4^{k-1} - y_3^{k-1})sinu_2 \tag{17}$$

$$y_4^k = y_3^k + \left(y_4^{k-1} - y_3^{k-1}\right)cosu_2 + \left(x_4^{k-1} - x_3^{k-1}\right)sinu_2 \tag{18}$$

where (x_2, y_2) denotes the coordinates of the known fixed-point K2; u_1 denotes the change in the joint angle between the cabin and the boom; u_2 is the sum of u_1 and the change of the joint angle between the arm and the boom. These state transition functions show that the estimates of K3 and K4 are not independent. Specifically, estimating K3 is based on the known fixed-point K2, and estimating K4 is based on the estimation of K3 at time *k*–1. Then, since the excavator model is nonlinear, these functions need to be linearized by first-order Taylor expansion, and the state transition matrix can be written as:

$$\mathbf{A} = \begin{bmatrix} cosu_1 & -sinu_1 & 0 & 0 \\ sinu_1 & cosu_1 & 0 & 0 \\ -cosu_2 & sinu_2 & cosu_2 & -sinu_2 \\ -sinu_2 & -cosu_2 & sinu_2 & -cosu_2 \end{bmatrix} \tag{19}$$

The process noise covariance is from the IMUs and given as:

$$\mathbf{Q} = I_4 \delta_{IMU}^2 \tag{20}$$

where δ_{IMU} is the variance in IMU noise. The measurement noise covariance is from the cameras and is given as:

$$\mathbf{R} = I_4 \delta_{CAM}^2 \tag{21}$$

where δ_{CAM} is the variance of camera noise. In addition, since the stereo vision module can directly provide the coordinates of K3 and K4 as the measurements, the observation matrix is given as:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{22}$$

So far, the trajectories of K3 and K4 in the projected 2D frame have been obtained by the proposed data fusion algorithm. These trajectories can be easily reconstructed from 2D to the 3D camera reference frame, which will be shown in Section 3.2.2. Additionally, to meet the needs of tracking multiple keypoints of an excavator in practice, there are two mechanisms specially designed in the proposed algorithm. The first mechanism is for synching the sampling rates of different sensors. In detail, the sampling rates of the IMUs are always much higher than that of cameras, so the IMU data needs to be integrated into the same sampling rates as the cameras to ensure consistent calculation. We therefore defined an adjustment parameter *n* in Eq. (23), which is equal to the integer portion of the ratio of the IMU's sampling frequency to the camera's sampling frequency. Before the competitive fusion, the *n* data provided by the IMUs are integrated from the camera statues *k*-1 to *k*, as the control input , to consistent the sampling rates of different sensors, as shown in Eq. (24).

$$n = [\frac{Sampling\ frequency\ of\ IMU}{Sampling\ frequency\ of\ camera}] \tag{23}$$

$$u_k = \sum_{i=n}^{1} IMU_i \tag{24}$$

The second mechanism is for monitoring outliers to enhance the robustness of the motion tracking system. There are two judgments for outliers: (1) If the differences between the measurement and the estimation exceed a preset threshold, the measurement will be accepted by the optimal result; (2) If measurements are lost, the estimations will be accepted by the optimal result. This mechanism allows users to adjust the fault tolerance of the algorithm based on their needs, improving the stability of the system in abnormal situations.

3.2.2. Tracking Other Keypoints of An Excavators Based on Complementary Data Fusion

This section determines the trajectories of the motions of all keypoints of an excavator in the world reference frame by complementarily fusing the optimal trajectories of partial keypoints (detailed in Section 3.2.1) with the motions of non-optimizable components (i.e., the cabin and the bucket) estimated by IMUs.

Due to the limited measurements provided by the cameras, the proposed multiple keypoints localization algorithm based on competitive data fusion can only obtain the trajectories of partial keypoints on the excavator (i.e., the arm joint (K3) and the bucket joint (K4)) in the camera reference frame. The end of the cabin (K1) and the boom joint (K2) are fixed points in the camera reference frame, which are only related to some physical information of the excavator (i.e., the length of the cabin, spatial distances between the boom joint and the camera). To describe the 3D full-body pose of the excavator, the location of the end of the bucket (K5) and the rotation of the cabin need to be estimated. Therefore, it is necessary to complementarily fuse data from IMUs attached to the bucket and the cabin to perform the measurements while the cameras cannot. First, as mentioned in Section 3.1.2, IMUs can independently estimate the joint angles between the bucket and the arm. Based on this relative angle, the locations of K5 can be easily appended to the incomplete excavator model in the camera reference frame by trigonometric functions. Afterwards, the IMU attached to the cabin complementarily provide the cabin rotating angle, which can help to transform the excavator motions from the camera frame to the world reference frame. Specifically, the transformation from the camera frame to the world frame required the orientation of cameras in the world frame. In our study, the cameras are mounted on the cabin so that the camera rotation is represented by the cabin rotating angles estimated by IMUs. This transformation acts on each keypoint of the excavator through a matrix T_{wc}, shown in Eq. (25).

$$T_{wc} = \begin{bmatrix} R_{wc} & t_{wc} \\ 0^T & 1 \end{bmatrix} \tag{25}$$

where the R_{wc} denotes the rotation of the camera frame relative to the pre-defined world frame, represented by a rotation matrix. This rotation is composed of the cabin rotating angle and a fixed rotation defined by the world frame in advance. t_{wc} denotes the position of the camera in the world frame, represented by a translation vector. Thus, based on complementary data fusion, the spatial coverage of the proposed competitive algorithm can be effectively extended and the full-body poses of the excavator are estimated in the world reference frame.

4. Experiments and Discussions

In this section, firstly, the EKF-based multiple keypoints localization algorithm developed in this study is applied on an excavator to test and evaluate its accuracy and robustness. Afterwards, based on the estimated locations of keypoints, full-body poses of the excavator are modeled to verify the feasibility of the proposed framework. More details are given in the subsections.

4.1. Experiment Setup

To fully prove the performance of the proposed framework in practical applications, the experiment was carried out on a real construction site using a real machine. Fig. 10 shows the devices used in the experiment. Image-based data acquisition was done using a fiducial marker (ArUco) attached onto the arm of the excavator and two RGB cameras embedded in mobile phones (OPPO Reno6), which formed a stereo module. The resolutions of the RGB cameras were 1280 x 720, and the frame rates were 30 frames per second (FPS). The cameras were installed on the front window of the excavator (model: FR65E2-H, make: LOVOL). In addition, IMU data was collected by the commercial IMU sensors LPMS-B2, equipped with embedded lithium batteries (3.7V@230mAh), which can work continuously for more than 6 hours, and with a sampling frequency of 100 Hz. These IMU sensors were non-invasively installed on the surface of each movable component (i.e., cabin, boom, arm, and bucket), which allows the sensors to be easily recharged, maintained, and replaced. The IMU is equipped with a Bluetooth transmitting and receiving module, which supports real-time data transmission (delay < 15ms), and the data was received and stored in a PC terminal within 20 meters. To validate the estimated pose of the excavator, another depth camera (RealSense D435i) was set on one side of the excavator to collect data as ground truth. By manually labeling and determining the positions of pre-defined keypoints in the depth camera coordinate system, the relative angles between adjacent components of the excavator were obtained. Then, according to the motions of each component pair and the structural relationship of the excavator, the locations of the pre-defined keypoints of the excavator were computed in the experimental coordinate system as ground truth. To ensure the reliability of the ground truth, the following methods were taken to reduce the potential noises of the measurements: (1) Depth information of the multiple points labeled near the pre-defined keypoints was averaged; (2) Two depth cameras simultaneously recorded the motions of the excavator, and their measurements were averaged; (3) The depth cameras were set close to the excavator about 2 meters.

(a) The excavator (b) The IMUs installed on the excavator (c)The stereo module and the fiducial marker

Fig. 10 Devices used in the on-site experiment

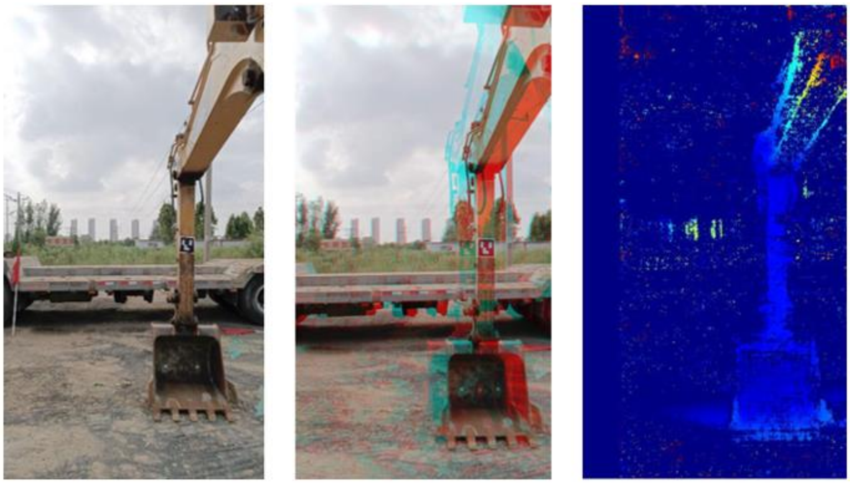
4.2. Performance Evaluation of the EKF-based Multiple Keypoints Localization Algorithm

The performance of the proposed algorithm as data-fusion-based keypoints localization of the excavator is evaluated and discussed on accuracy and robustness in two cases: (1) independent motion and (2) continuous motion.

In the case of independent motion, the components of the excavator are operated, including the lifting and lowering of the boom and arm, independently respectively. This case focuses on verifying the performances of the proposed algorithm in tracking a single keypoint of an excavator so only the point directly affected by the independent motion is concerned in this case. Specifically, when the boom moves, the performance (i.e., accuracy and robustness) of tracking the arm joint (K3) is evaluated; When the arm moves, the performance of tracking the bucket joint (K4) is evaluated. The boom trial involves four repeated cycles of boom motions and evaluates 2100 sets of IMU data and 630 independent measurements from the camera. The arm trial includes four repeated cycles of arm motions, and 2185 sets of IMU data and 656 independent measurements from the camera are evaluated. The data contains all the motion modes of the components, so it is diverse.

Pose information of the excavator estimated by IMUs is computed using an existing method which has been evaluated in [12] in detail. Raw data is collected by IMUs attached on different movable components of the excavator and processed to estimate the orientation of each component using the method explained in Section 3.1.2. The static initial pose of the excavator required for IMU-based pose estimation is provided by the stereo vision module. These orientations of movable components are calculated into the pose information from IMUs which is required by the proposed data-fusion-based keypoints localization algorithm.

The image data describing the current pose of the excavator is captured by a stereo vision module composed of two RGB cameras installed on the cabin. The baseline between the cameras is 100 mm to ensure a relatively stable acquisition of the depth information of the feature point when the arm of the excavator is away from the cameras. After camera calibration, each standardized images pair is generated a stereo anaglyph and a disparity map, and the depth information of each recognizable point is reconstructed on the left view. Fig. 11 illustrates an example of the original image, the stereo anaglyph, and the disparity map in the experiment. Then, the contour of the fiducial marker is identified on the corresponding left view, and the pre-defined feature point as the centroid of the marker is determined on the image based on the contour, as introduced in Section 3.1.3. The coordinates of the feature point in the camera reference frame are obtained by retrieving the depth information of the centroid. Combined with the coordinates of the feature point and the physical parameters of the excavator, the trajectory of K3 and K4 are obtained in the camera frame by the proposed geometric decoder. Table 1 shows the physical parameters of the excavator in the image-based onboard motion tracking. The trajectories of K3 and K4 provide a direct observation for the locations of keypoints which are the inputs of the data-fusion-based localization method from cameras.



(a) The original image (b) The stereo anaglyph (c) The disparity map

Fig. 11 An example of the original image, the stereo anaglyph, and the disparity map

Table 1 Physical parameters of the excavator used in the image-based on-board motion tracking

Physical parameters	Length (mm)

L1: Length of the boom	3100
L2: Distance from the arm joint to the centroid of the marker	600
L3：Length of the arm	1500
L4: Horizontal distance from the stereo module to the boom joint	365
L5: Vertical distance from the stereo module to the boom joint	270
L6: Depth from the stereo module to the boom joint	340

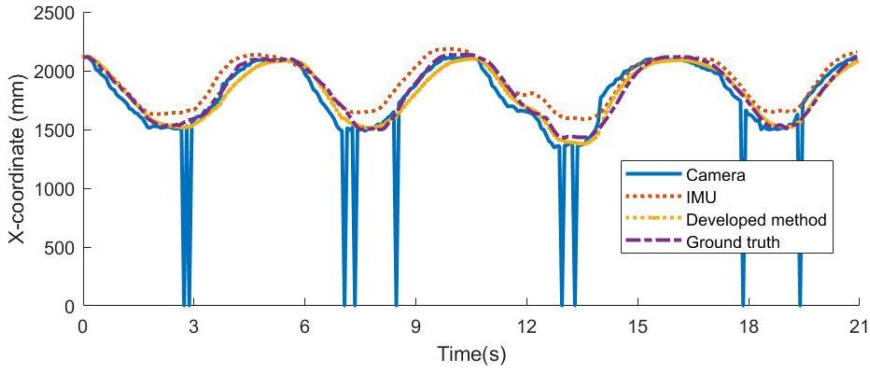
After synchronizing the first moving point to align different data on the timeline, the pose information contributed from the IMUs and cameras is merged and inputted into the keypoint localization algorithm. The parameters used for the algorithm tunning are listed in Table 2. Since the articulated parts of the excavator are coplanar in the camera frame, the performance of tracking K3 and K4 is evaluated on the projected 2D frame. In this study, the root mean squared error (RMSE) is used to represent the average errors of the estimated keypoint location, as it is common to use RMSE to measure the differences between estimated values and ground truths. Table 3 shows the results and their RMSEs in the case of independent motions.

Table 2 Parameters for the proposed algorithm tunning

Variables	Meanings
Sampling interval of IMU sensors	100HZ
Sampling interval between image frames	30 FPS
Noise variance of IMU sensors, δ_{IMU}	0.1
Noise variance of cameras, δ_{CAM}	0.59

Table 3 Results of the independent motion case in *x*- and *y*- axes

Components/Keypoint	Results	RMSEs(mm)

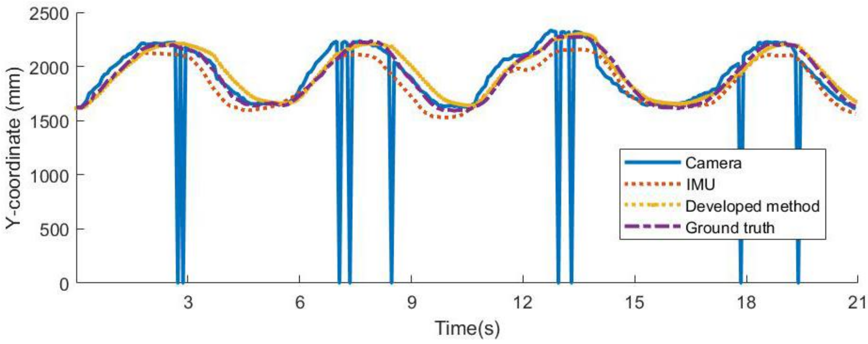


Camera_X= 80.41

IMU_X= 96.15

Optimal_X= 45.59

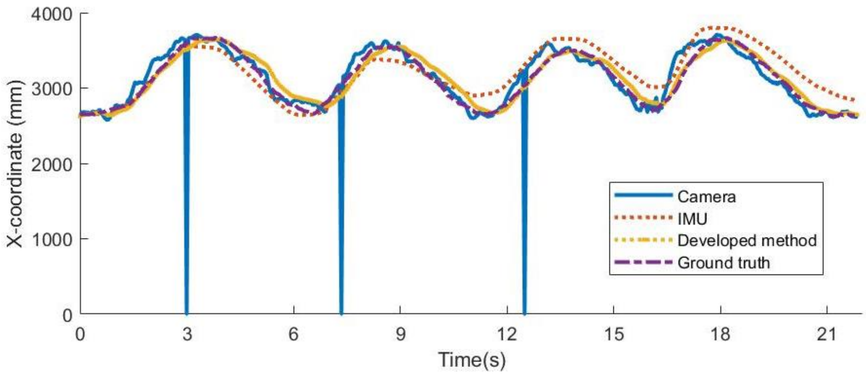
Boom/K3



Camera_Y= 75.64

IMU_Y= 89.28

Optimal_Y= 55.45

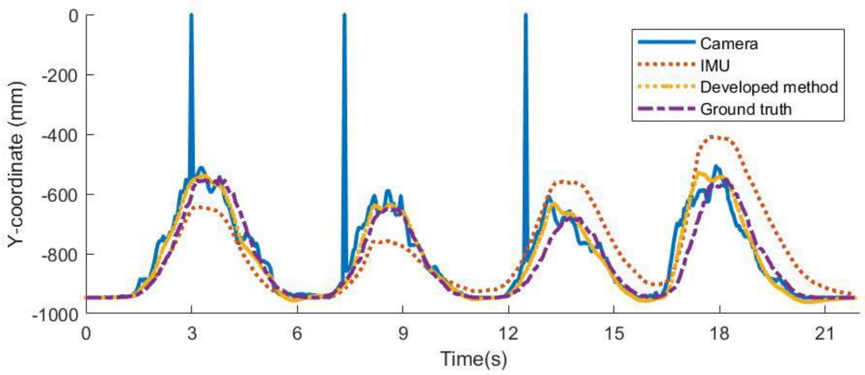


Camera_X=108.91

IMU_X= 147.43

Optimal_X= 84.09

Arm/K4



Camera_Y=51.73

IMU_Y= 87.63

Optimal_Y= 40.39

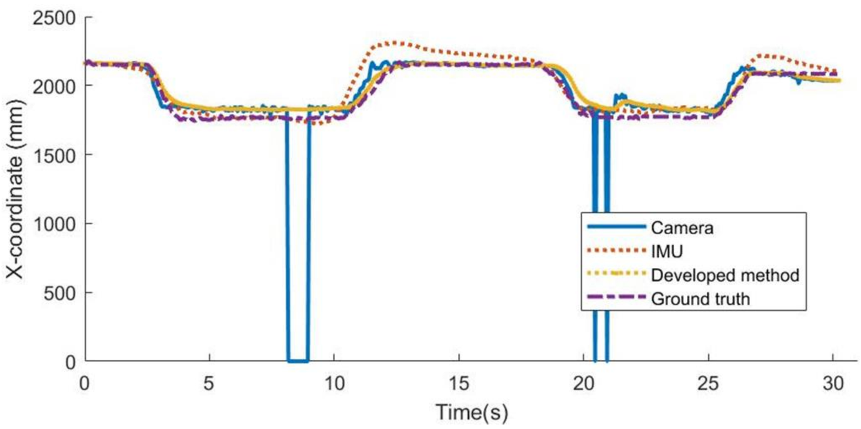
The trajectories of K3 and K4 are illustrated on the *x*-axis and *y*-axis respectively. Each figure includes four curves: the trajectory as directly observed by the stereo vision module; the trajectory estimated by the IMUs, the optimized trajectory estimated by the developed algorithm, and the ground truth. Through comparing the curves, the robustness and accuracy of the developed algorithm are verified and discussed. Four distinct cycles, corresponding to the four repeated independent motions in each trial, can be observed in each curve. The amplitudes of these curves are consistent with the normal operation of an excavator. In terms of robustness, the results show that there are some outliers or zeros during the process of tracking keypoints using cameras. Such points represent the loss or large deviation of the image data captured at any given time. It is speculated that these noises are caused by sparse disparity maps or unrecognized makers due to environmental changes. In addition, the trajectories obtained by the cameras have obvious noise coming from the vibrations of the moving component and the unavoidable slight displacement of the cameras with the operation of the excavator. For IMU sensors, the results shown in Table 3 are obtained by integrating the IMU data directly. Obvious biases are observed in the trajectories, which exceed 1 degree over 20 seconds and increase with time. It is speculated that these biases are caused by accumulating drifts of gyroscopes. In general, when relying on homogeneous sensors, especially the cameras, the keypoint localization system shows obvious instability, which may cause great deviation in the results or even failure of the system. In contrast, the trajectories estimated by the developed method are smooth and stable, which compensates for the missing observations of cameras and optimizes significant bias through the data from another source. Therefore, the experimental results demonstrate that the proposed sensor-fusion-based keypoints localization method is more robust than the method using homogeneous sensors, in independent motion tracking. It means that the developed method is less susceptible to extreme cases with data loss and obvious biases. To further investigate the accuracy, Table 4 shows overall results and average errors of the different methods on the keypoints localization which are calculated based on the estimated results in the x-axis and y-axis provided in Table 3. According to the RMSEs listed in Tables 3 and 4, the difference between the estimated trajectory obtained by the proposed method and the ground truth is in the range of 40 to 84 mm, and the average errors are 73.76mm. Compared with the average errors of 115.48mm based on camera observation and 151.36mm based on IMU estimation, it is found that the proposed method effectively improved the accuracy of keypoint localization. In addition, it is also observed that the errors of K4 localization are always slightly greater than K3, because the vibration of the arm caused by inertia is more obvious than that of the boom when the components of the excavator are moving. In summary, the proposed sensor-fusion-based keypoints localization algorithm has better robustness and accuracy than the direct visual observations or IMU-based estimation, in the case of independent motion tracking.

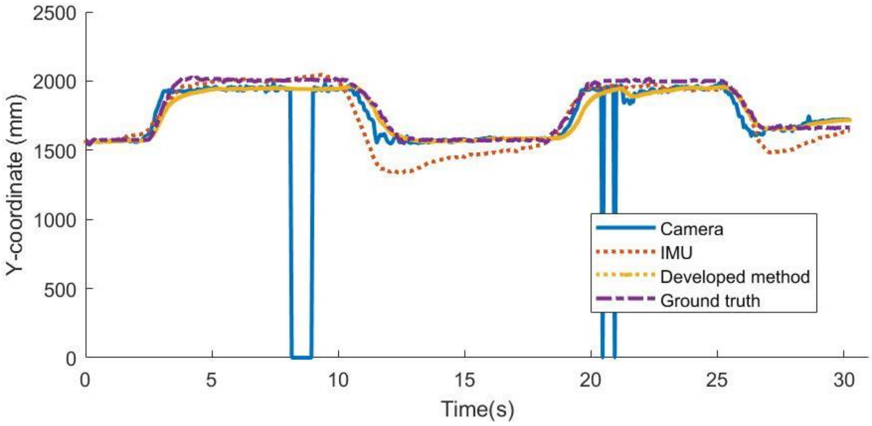
Table 4 Spatial RMSEs and average errors of the independent motion

Methods	RMSEs_K3(mm)	RMSEs_K4(mm)	Average(mm)
Camera	110.39	120.57	115.48
IMU	131.21	171.51	151.36
Developed method	71.78	75.73	73.76

To further investigate the effectiveness of the proposed method in the working states of the excavator with continuous motions, the second case focuses on using the proposed algorithm to track multiple keypoints (i.e., K3 and K4) simultaneously when the excavator digs and dumps. Each trial involves two repeated full working cycles of digging and dumping. In each cycle, multiple components of the excavator moved continuously, including the left and right rotation of the cabin, the up and down motion of the boom, the arm, and the bucket respectively. In this case, 3024 sets of IMU data and 907 independent measurements from the camera were respectively evaluated for each keypoint. The data contains all the motion modes of digging and dumping in practical, so it is diverse. Table 5 shows the results and their RMSEs in the case of continuous motions.

Table 5 Results of the continuous motion case in the *x*- and *y*- axes

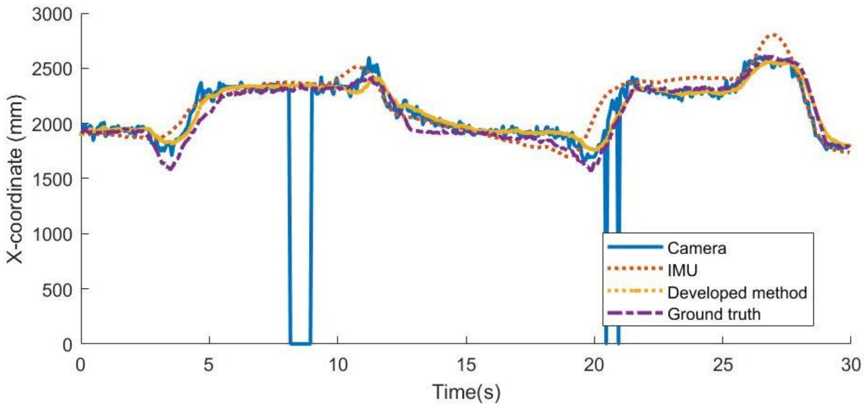
Keypoints	Results	RMSEs
K3		Camera_X= 58.88 IMU_X= 96.15 Optimal_X= 54.89



Camera_Y= 62.45

IMU_Y= 111.76

Optimal_Y= 56.88

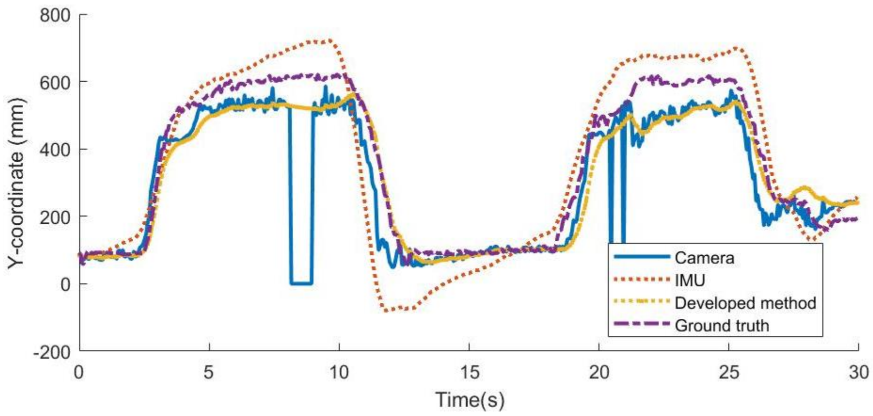


Camera_X= 111.47

IMU_X=166.50

Optimal_X= 77.96

K4



Camera_Y= 106.08

IMU_Y= 125.94

Optimal_Y= 66.21

From the figures of the trajectories in the x-axis and y-axis, two clear cycles can be observed in the results achieved by different tracking methods, corresponding to the two continuous work cycles of digging and dumping. From the perspective of robustness, similar to the first case, the keypoints trajectories obtained by the cameras are unstable, which can be obviously observed in the figures. Besides the outliers discussed in the first case, some continuous data losses were observed with an interval of 5 to 10 s. A possible reason is that as the cabin rotates, the changes of illumination render the fiducial marker unrecognizable. The instability of the IMU-based location estimation manifested itself in the obvious drift caused by accumulated biases, which were observed on the trajectories computed through the direct integration of IMU data. Compared with the keypoints' trajectories achieved by cameras and IMUs, the proposed method based on sensor fusion optimizes the estimated results by data from different sources, leading to smoother and more stable trajectories. Especially in the interval where the camera observations are missing, the proposed algorithm can still estimate the motion of the excavator by data from the other source – IMU, which means that the proposed algorithm degenerates into an IMU-only method but keeps the basic survivability and stability of the pose estimation system. Thus, the proposed method can effectively improve the robustness of multiple keypoints localization for excavators on construction sites. Table 6 shows the overall results and average errors of the continuous motion, which is calculated based on the RMSEs in the x- and y-axes provided in Table 5. Based on Table 6, the differences between the trajectories estimated by the proposed method and the ground truths are in the range of 54.89 to 102.28 mm with its average as 90.66mm. This result is less than the errors of camera observation as 119.85mm and IMU estimation 178.10mm. Therefore, it is proved that in the case of continuous motions, the proposed method can improve the robustness and accuracy of tracking multiple keypoints of excavators on construction sites.

Fig. 12 illustrates the RMSEs of the estimated results based on the proposed method from the cases of independent motions and continuous motions. There is no significant difference in the trends of RMSEs in the two cases, and the total average numerical error for tracking the multiple keypoints of the excavator is 82.21 mm in value. In order to intuitively show the improved accuracy of the proposed algorithm, the average percent error (as a percentage of the total traveled distance) [38] is used to evaluate the errors of different methods. According to the above experimental results, the average percent error of the proposed algorithm accounts for 1.21% of the total traveling length (29372 mm computed by ground truth), which is lower than the error of the IMU-based approach at 2.38% and of the camera-based approach at 1.65%. Besides, in Fig. 12, it is observed that the errors of the second case are slightly larger than the of the first case. Here, two reasonable inferences are provided about this phenomenon: (1) When multiple keypoints of an excavator were tracked simultaneously, the estimation uncertainties of the previous keypoint were inherited by the following one. It means that the uncertainties were accumulated in K4, resulting in a relatively large deviation in the K4 localization; (2) In the case of continuous motions, the rotating cabin and the moving bucket brought more unmodeled vibrations for the keypoints localization. Especially, if the machine stopped emergently, the strong swing of the articulated parts of the excavator on both the x- and y-axes caused by inertia affects the overall estimation accuracy. In summary, compared with the existing IMU-based pose estimation method and image-based motion tracking method for excavators, this experiment verifies that the proposed sensor-fusion-based algorithm can effectively improve the robustness and accuracy of keypoints localization for excavators, for estimating both single keypoint in the independent motions and multiple keypoints in continuous motions.

Table 6 Spatial RMSEs and average errors of the continuous motion

Methods	RMSEs_K3(mm)	RMSEs_K4(mm)	Average(mm)
Camera	85.83	153.87	119.85
IMU	147.43	208.77	178.10
Developed method	79.04	102.28	90.66

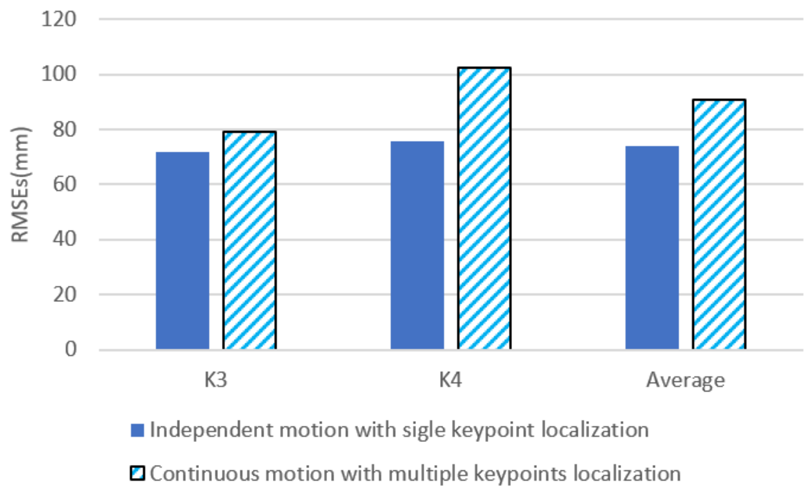


Fig. 12 Comparison of the RMSEs of estimated results from case 1 and case 2

Considering the changing implementation conditions on sites, the conclusion drawn from the experiment needs to be further discussed. Firstly, to investigate the influence of different excavator models on the performance of the proposed algorithm, a large excavator (model: ZAXIS 240, make: HITACHI) was used to repeat the continuous motions of digging and dumping in the third case and the estimation results were compared with the medium-sized excavator in the second case. Table 7 lists the specifications of the large excavator. To render the acquisition of depth information relatively stable, the baseline of the stereo vision module was increased to 150 mm. Other hardware configurations were the same as in the previous experiment. Fig. 13 illustrates the large excavator and the installation of the onboard devices. In this case, 3020 sets of IMU data and 906 independent measurements from the camera were respectively evaluated for each keypoint. Table 8 shows the estimated results and their RMSEs in the case of the large excavator.

Table 7 Physical parameters of the large excavator

Physical parameters	Length (mm)
L1: Length of the boom	4560
L2: Distance from the arm joint to the centroid of the marker	1000
L3： Length of the arm	2900
L4: Horizontal distance from the stereo module to the boom joint	790
L5: Vertical distance from the stereo module to the boom joint	510
L6: Depth from the stereo module to the boom joint	620

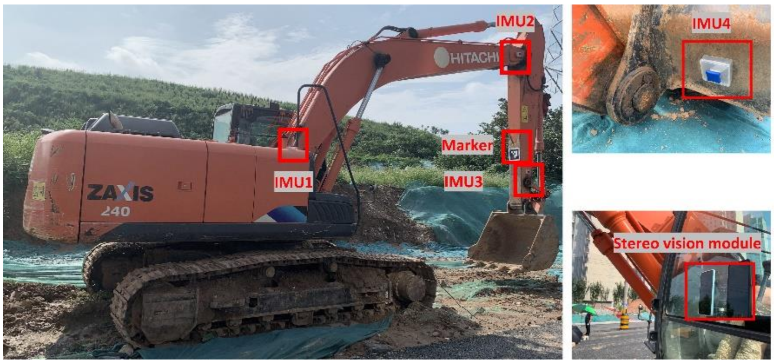
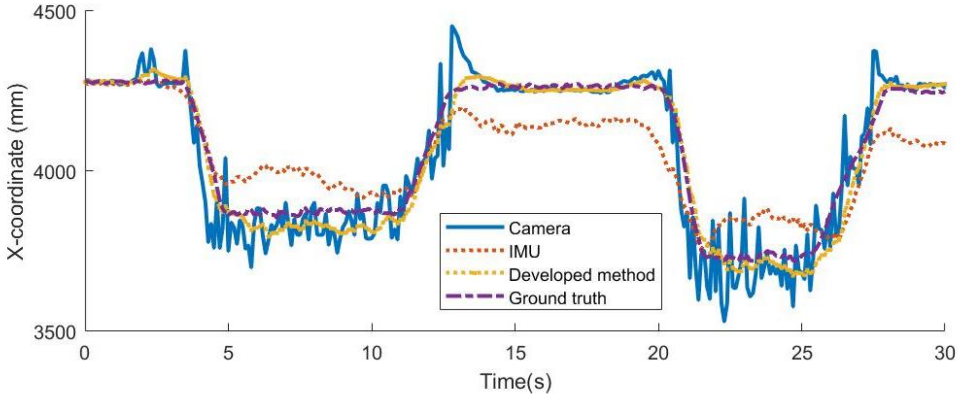


Fig. 13 Large excavator and the installation of the onboard devices

Table 8 Results of using the large excavator in the continuous motion case

Keypoints	Results	RMSEs

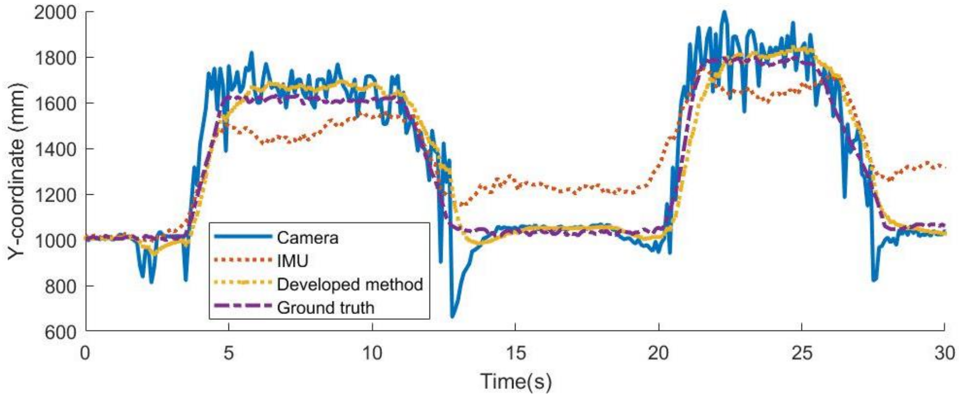


Camera_X= 77.09

IMU_X= 104.26

Optimal_X= 61.34

K3

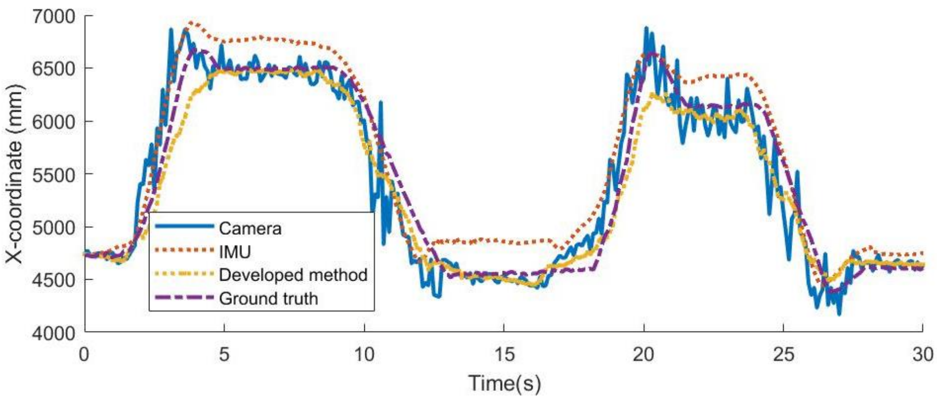


Camera_Y= 108.76

IMU_Y= 155.04

Optimal_Y= 63.13

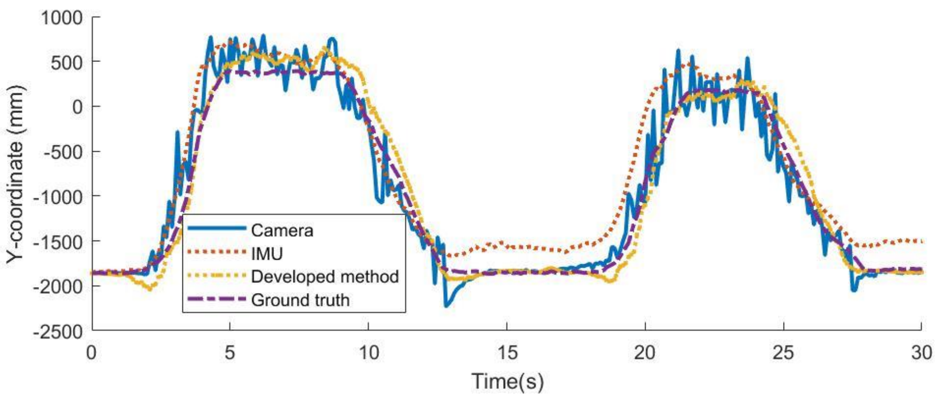
K4



Camera_X= 257.43

IMU_X=255.67

Optimal_X= 179.88



Camera_Y= 225.26

IMU_Y= 296.58

Optimal_Y= 135.90

In terms of robustness, it was observed that the proposed algorithm provides smoother trajectories with better stability in keypoint tracking for the large excavator, compared to the IMU-only and camera-only methods. In terms of accuracy, compared to the results of the medium-sized excavator shown in Table 6, the RMSEs of the large-sized excavator in Table 8 have a slight increase on all axes. However, according to Table 8, the average percent error of applying the proposed algorithm to the large excavator accounts for 1.11% (total travel distance: 74649 mm), which is lower than the error of the camera-only method at 1.53% and the error of the IMU-only method at 2.07%, and close to the average percent error of using the medium-sized excavator at 1.20%. Therefore, the proposed algorithm still obtains the smallest average error of keypoint tracking and optimized accuracy performance compared to the IMU-only and camera-only methods. This result can be supported by theoretical analysis: the different excavator models would not import new uncertainty into the proposed algorithm, thus the effect of optimization on the accuracy and robustness of pose estimation is not affected by different sizes of excavators. In summary, though the numerical accuracy may change with the size of the excavator, it can be generalized that compared with tacking poses using homogenous sensors, the proposed algorithm can improve the accuracy and robustness of the pose estimation, regardless of different excavator models.

In addition to the specification of excavators, varying visual conditions (e.g., changing backgrounds and lack of illumination in bad weather) also need to be considered in practice. First, to avoid the impact of background changes on the proposed algorithm, the image-based motion tracking in Section 3.1.3 uses the binary square fiducial marker

– ArUco and its pre-defined library, including a wide black border and an inner binary matrix which is uniquely identified based on the library. As a result, ArUco markers can be robustly identified regardless of the changing background [35], which ensures that the proposed framework is able to consistently acquire visual observations in different backgrounds. It has been also verified by the third case with the large excavator. According to Table 8, although the third case changes the background and visual conditions compared to the second case using the medium excavator, the observations of the large excavator's motions were steadily obtained by identifying the ArUco marker attached to its arm. Therefore, it is concluded that the proposed framework is not affected by the changing background.

Additionally, bad weather (insufficient illumination) may make the fiducial marker unrecognizable resulting in losing observations of the cameras. In this case, according to the principle of EKF [37], the proposed algorithm degenerates into the IMU-based pose estimation [12]. This degenerated situation has been verified in the 5-10s interval in case 2. Although the degraded algorithm loses the accuracy improvement brought by competitive data fusion, it keeps the basic survivability and stability of the pose estimation system, which is an advantage of the proposed algorithm based on sensor fusion. There are many studies dedicated to improving the quality of visualization in bad illumination (e.g., Zheng et al. [39]), which can facilitate the proposed algorithm to maintain accurate and stable estimation in bad weather.

4.3. Full-body Pose Modeling of an Excavator

To verify the feasibility of the proposed pose estimation framework for excavators, this section continuously models the full-body pose of the excavator using MATLAB 2020b based on the optimal trajectories of the K3 and K4 in the camera reference frame and data collected from the digging and dumping medium-sized excavator on construction sites.

After obtaining the optimal trajectories of the K3 and K4 using the proposed multiple keypoints localization method, the full trajectories of other excavator keypoints need to be obtained, i.e., the end of the cabin (K1), the boom joint (K2), and the end point of the bucket (K5) in the camera reference frame. K1 and K2 are fixed points only related to the physical parameters of the excavator. The determination of K5 requires the joint angle between the bucket and the arm estimated by the IMUs installed on the bucket and the arm, which is illustrated in Fig. 14 as Theta 2. The physical parameters required in the determination of the K1, K2, and K5 are listed in Table 9. Then, to reconstruct all the keypoints from 2D to 3D space, it is necessary to obtain the transformation matrix in a pre-defined world reference frame. In practice, the world reference frame is flexibly selected according to the user's needs, so it usually different from the camera reference frame and need to be transformed. To show the process, the world reference frame is defined as a right-hand system, where the y-axis is the rotation axis of the excavator, the z-axis points to the initial optical axis of the cameras, and the origin is located on the ground. Thus, the transformation calculation is shown in Eq. (31).

$$\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} cos\theta_t & 0 & sin\theta_t & -x_s \\ 0 & 1 & 0 & -y_s \\ -sin\theta_t & 0 & cos\theta_t & -z_s \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{31}$$

where θ_t denotes the cabin's angle of rotation at time t (shown in Fig. 14 as Theta 1); (x_w, y_w, z_w) denote the coordinates of the keypoint in the world reference frame; (x_c, y_c, z_c) are the coordinates of the keypoint in the camera reference frame; (x_s, y_s, z_s) are the coordinates of the cameras in the world reference frame, which are listed in Table 9. Then, the full-body poses of the excavator are represented by the motion trajectories of all keypoints determined in the pre-defined, world reference frame. Fig. 15 shows two examples of the full-body pose modeling of the excavator at two time slots. Additionally, considering the requirement of operational safety monitoring on response time in practice, the proposed framework was conducted a timing-test on the laptop (model name: Lenovo Legion Y7000P2021, CPU: i7-11800H, GPU: GeForce RTX 3050Ti). The average response time of full-body pose estimation at each time slot based on the proposed framework is 0.038s, i.e., such data inference stage will cause a negligible delay in practice to track the pose of an excavator. Hence, this proposed framework can meet the needs of real-time data processing of operational safety monitoring in practice. It is proved that the proposed full-body pose estimation framework of excavators is feasible and reliable on construction sites.

Table 9 Physical parameters of the excavator used in the 3D modeling

Physical parameters	Length (mm)
Length of the cabin	1950

Length of the bucket	890
x-coordinate of the camera (x_s)	100
y-coordinate of the camera (y_s)	975
z-coordinate of the camera (z_s)	100

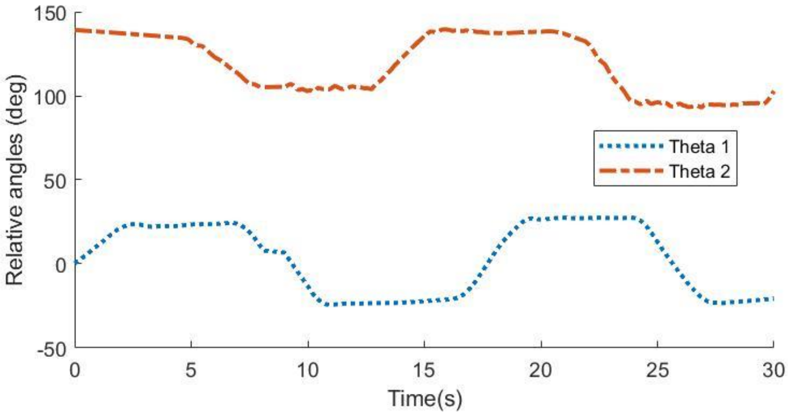


Fig. 14 Trajectories of the cabin's angle of rotation (Theta 1) and the angle between the arm and the bucket (Theta 2)

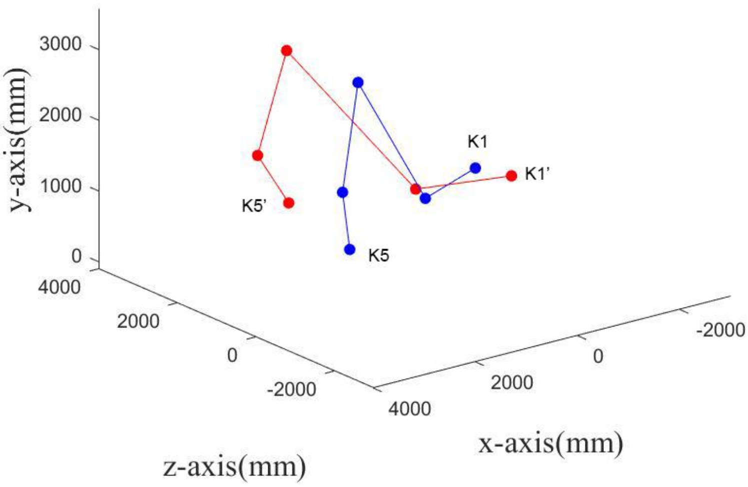


Fig. 15 Examples of the full-body pose modeling of the excavator in the world reference frame

5. Conclusions

This study proposes a full-body pose estimation framework for excavators that uses data fusion of multiple onboard sensors. In this framework, a non-invasive onboard visual-inertial system is developed to track the excavator motions on construction sites. Then, through competitive and complementary data fusion, the keypoints describing full-body poses of the excavator are tracked in 3D space. In particular, an EKF-based multiple keypoint localization algorithm is developed to merge the pose information obtained from IMUs and cameras and optimize estimations of multiple keypoints of excavators simultaneously. A real case study verified that the proposed multiple keypoint localization algorithm effectively improved the robustness and accuracy of tracking pre-defined excavator keypoints. The experimental results show that, compared with using homogeneous sensors, the trajectories estimated by the proposed algorithm are smoother and more stable, and it has stronger survivability in complex situations on construction sites (e.g., data loss and strong vibration). The average RMSEs of the tested medium excavator between the estimated results based on the proposed algorithm and the ground truth is 82 mm in value. The average percent error of the proposed algorithm accounts for 1.21% of the total travelled distance, which is lower than 2.38% for the IMU-based method and 1.65% for the camera-based method. The proposed framework based on data fusion of multiple onboard sensors provides the theoretical basis for developing an accurate and robust 3D full-body pose estimation of excavators on real construction sites to monitor the motions of machinery in real-time and improve the operational safety.

The limitation of the proposed framework is that the lack of the specific noise model of the excavator working on construction sites limits the accuracy of the proposed sensor-fusion-based multiple keypoint localization algorithm. In future works, based on further analysis of the error sources of the visual-inertial sensor system, the noises of a working excavator, such as strong vibration caused by inertia and environmental interferences, can be modeled, which will improve the accuracy of the proposed algorithm.

References

[1] Labour Department, "Occupational safety and health statistics bulletin," 2021, issue 21.https://www.labour.gov.hk/common/osh/pdf/Bulletin2020_issue21_en.pdf, (accessed: June 7 2022)

[2] Ministry of Housing and Urban-Rural Development of the People's Republic of China, "Report of Safety Accidents in China's Building Construction Activities in 2019," 2020.https://www.mohurd.gov.cn/gongkai/fdzdgknr/tzgg/202006/20200624_2

46031.html, (accessed: August 20 2022)

[3] U.S. Bureau of Labor Statistics, "Census of fatal occupational injuries," 2021.https://www.bls.gov/news.release/cfoi.htm, (accessed: June 13 2022)

[4] Labour Department, "Occupational safety and health statistics 2020," 2021.https://www.labour.gov.hk/common/osh/pdf/archive/statistics/OSH_Statistics_2020_en.pdf, (accessed: June 7 2022)

[5] Occupational Safety and Health Administration, "Top four construction hazards," 2021.https://www.osha.gov/sites/default/files/publications/construction_hazards_qc.pdf, (accessed: June 17 2022)

[6] S. Xu, J. Wang, W. Shou, T. Ngo, A.-M. Sadick, and X. Wang, "Computer vision techniques in construction: A critical review," *Archives of Computational Methods in Engineering,* pp. 1-15, 2020.https://doi.org/10.1007/s11831-020-09504-3

[7] Schiffbauer and W. H, "An active proximity warning system for surface and underground mining applications," *Mining Engineering,* vol. 54, pp. 40-48, 2002.https://www.cdc.gov/niosh/mining/works/coversheet609.html (accessed: January 1 2022)

[8] H. Wu *et al.*, "A location based service approach for collision warning systems in concrete dam construction," *Safety Science,* vol. 51, no. 1, pp. 338-346, 2013.https://doi.org/10.1016/j.ssci.2012.08.006

[9] Y. Kim, J. Baek, and Y. Choi, "Smart helmet-based personnel proximity warning system for improving underground mine safety," *Applied Sciences,* vol. 11, no. 10, p. 4342, 2021.https://doi.org/10.3390/app11104342

[10] H. Luo, M. Wang, P. K.-Y. Wong, and J. C. Cheng, "Full body pose estimation of construction equipment using computer vision and deep learning techniques," *Automation in Construction,* vol. 110, p. 103016, 2020.https://doi.org/10.1016/j.autcon.2019.103016

[11] J. Zhao, Y. Hu, and M. Tian, "Pose estimation of excavator manipulator based on monocular vision marker system," *Sensors,* vol. 21, no. 13, p. 4478, 2021.https://doi.org/10.3390/s21134478

[12] J. Tang, H. Luo, W. Chen, P. K.-Y. Wong, and J. C. Cheng, "IMU-based full-body pose estimation for construction machines using kinematics modeling," *Automation in Construction,* vol. 138, p. 104217,

2022.https://doi.org/10.1016/j.autcon.2022.104217

[13] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 3400-3407: IEEE.https://doi.org/10.1109/ICRA.2011.5979561

[14] E. R. Azar, C. Feng, and V. R. Kamat, "Feasibility of in-plane articulation monitoring of excavator arm using planar marker tracking," *Journal of Information Technology in Construction (ITcon),* vol. 20, no. 15, pp. 213-229, 2015.https://www.itcon.org/2015/15, (accessed: June 13 2020)

[15] M. M. Soltani, Z. Zhu, and A. Hammad, "Skeleton estimation of excavator by detecting its parts," *Automation in Construction,* vol. 82, pp. 1-15, 2017.https://doi.org/10.1016/j.autcon.2017.06.023

[16] M. M. Soltani, Z. Zhu, and A. Hammad, "Framework for location data fusion and pose estimation of excavators using stereo vision," *Journal of Computing in Civil Engineering,* vol. 32, no. 6, p. 04018045, 2018.https://doi.org/10.1061/(ASCE)CP.1943-5487.0000783

[17] T. Phillips, "Determining and verifying object pose from LiDAR measurements to support the perception needs of an autonomous excavator," School of Mechanical and Mining Engineering, The University of Queensland, 2016.https://doi.org/10.14264/uql.2016.787

[18] T. G. Phillips and P. R. McAree, "An evidence-based approach to object pose estimation from LiDAR measurements in challenging environments," *Journal of Field Robotics,* vol. 35, no. 6, pp. 921-936, 2018.https://doi.org/10.1002/rob.21788

[19] C. Zhang, A. Hammad, and S. Rodriguez, "Crane pose estimation using UWB real-time location system," *Journal of Computing in Civil Engineering,* vol. 26, no. 5, pp. 625-637, 2012.https://doi.org/10.1061/(ASCE)CP.1943-5487.0000172

[20] S. Lee, M.-S. Kang, D.-S. Shin, and C.-S. Han, "Estimation with applications to dynamic status of an excavator without renovation," in *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, 2012. https://doi.org/10.22260/ISARC2012/0093

[21] S. Talmaki and V. R. Kamat, "Real-time hybrid virtuality for prevention of excavation related utility strikes," *Journal of Computing in Civil Engineering,*

vol. 28, no. 3, p. 04014001, 2014.https://doi.org/10.1061/(ASCE)CP.1943-5487.0000269

[22] F. A. Bender, S. Göltz, T. Bräunl, and O. Sawodny, "Modeling and offset-free model predictive control of a hydraulic mini excavator," *IEEE Transactions on Automation Science and Engineering* vol. 14, no. 4, pp. 1682-1694, 2017.https://doi.org/10.1109/TASE.2017.2700407

[23] Z. Péntek, T. Hiller, T. Liewald, B. Kuhlmann, and A. Czmerk, "IMU-based mounting parameter estimation on construction vehicles," in *Proceedings of the 2017 DGON Inertial Sensors and Systems (ISS)*, 2017, pp. 1-14.https://doi.org/10.1109/InertialSensors.2017.8171504

[24] O. J. Woodman, *An introduction to inertial navigation*. University of Cambridge, Computer Laboratory, 2007.https://doi.org/10.48456/tr-696

[25] W. Elmenreich, "An introduction to sensor fusion," Vienna University of Technology, Austria2002, vol. 502.https://www.researchgate.net/profile/Wilfried-Elmenreich/publication/267771481_An_Introduction_to_Sensor_Fusion/links/55d2e45908ae0a3417222dd9/An-Introduction-to-Sensor-Fusion.pdf (accessed: January 22 2022)

[26] S.-H. Kim *et al.*, "Development of bulldozer sensor system for estimating the position of blade cutting edge," *Automation in Construction,* vol. 106, p. 102890, 2019.https://doi.org/10.1016/j.autcon.2019.102890

[27] S. Moberg, J. Öhr, and S. Gunnarsson, "A benchmark problem for robust control of a multivariable nonlinear flexible manipulator," *IFAC Proceedings Volumes,* vol. 41, no. 2, pp. 1206-1211, 2008.https://doi.org/10.3182/20080706-5-KR-1001.00208

[28] P. Axelsson, M. Norrlöf, E. Wernholt, and F. Gustafsson, "Extended kalman filter applied to industrial manipulators," in *Proceedings of Reglermöte 2010*, 2010.http://www.diva-portal.org/smash/record.jsf?dswid=4916&pid=diva2%3A606591 (accessed: January 3 2022)

[29] T. M.Ruff, "Recommendations for evaluating and implementing proximity warning systems onsurface mining equipment," in "Report of Investigations (National Institute for Occupational Safety and Health) " 2007.https://stacks.cdc.gov/view/cdc/8494 (accessed: January 2 2022)

[30] B. Liu, F. Zhang, and X. Qu, "A method for improving the pose accuracy of a robot manipulator based on multi-sensor combined measurement and data fusion," *Sensors,* vol. 15, no. 4, pp. 7933-7952, 2015.https://doi.org/10.3390/s150407933

[31] B. Ubezio, S. Sharma, G. Van der Meer, and M. Taragna, "Kalman filter based sensor fusion for a mobile manipulator," in *Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2019, vol. 59230, p. V05AT07A043: American Society of Mechanical Engineers.https://doi.org/10.1115/DETC2019-97241

[32] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proceedings of the IEEE International Conference on Rehabilitation Robotics*, 2011, pp. 1-7.https://doi.org/10.1109/ICORR.2011.5975346

[33] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, no. 11, pp. 1330-1334, 2000.https://doi.org/10.1109/34.888718

[34] R. I. Hartley, "Euclidean reconstruction from uncalibrated views," in *Joint European-US Workshop on Applications of Invariance in Computer Vision*, 1993, pp. 235-256: Springer.https://doi.org/10.1007/3-540-58240-1_13

[35] R. Munoz-Salinas, "Aruco: a minimal library for augmented reality applications based on opencv," in "Universidad de Córdoba," 2012, vol. 386.https://www.uco.es/investiga/grupos/ava/node/26 (accessed: February 2 2022)

[36] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.https://doi.org/10.1108/k.2001.30.9_10.1333.2

[37] G. A. Terejanu, "Extended kalman filter tutorial," University at Buffalo2008.https://www.cse.sc.edu/~terejanu/files/tutorialEKF.pdf (accessed: March 8 2022)

[38] M. Ibrahim and O. Moselhi, "Inertial measurement unit based indoor localization for construction applications," *Automation in Construction,* vol. 71, pp. 13-20, 2016.https://doi.org/10.1016/j.autcon.2016.05.006

[39] C. Zheng, D. Shi, and W. Shi, "Adaptive unfolding total variation network for low-light image enhancement," in *Proceedings of the IEEE/CVF International*

Conference on Computer Vision, 2021, pp. 4439-4448.https://doi.org/10.1109/iccv48922.2021.00440