

Application of Simultaneous Localization and Mapping for Large-scale Manipulators in Unknown Environments

Petri Mäkinen, Mohammad M. Aref and Jouni Mattila
Unit of Automation Technology and Mechanical Engineering
Tampere University
Tampere, Finland
Emails: petri.makinen@tuni.fi,
m.aref@ieee.org,
jouni.mattila@tuni.fi

Sirpa Launis
Sandvik Mining and Construction
Tampere, Finland
Email: sirpa.launis@sandvik.com

Abstract—In this paper, we study the application of simultaneous localization and mapping (SLAM) for estimating the tool center point (TCP) 6 degrees-of-freedom (DOF) pose of a large-scale hydraulic manipulator without a priori knowledge of the environment. We attach a stereo camera near the TCP of the manipulator and perform SLAM by utilizing the open source version of ORB-SLAM2. In offline experiments, the camera frame and the TCP frame are extrinsically calibrated using an iterative closest point search to match a point cloud of poses from the SLAM module with a point cloud of ground-truth TCP poses, which are obtained from joint encoder measurements along with a kinematic model of the manipulator. The estimated TCP trajectory provided by the SLAM is then compared to the ground-truth TCP trajectory. These preliminary experiments show that a pure visual SLAM algorithm can perform reasonably well in this application scenario. Limitations and future work are also discussed.

Index Terms—manipulators, simultaneous localization and mapping, position measurement, rotation measurement

I. INTRODUCTION

Autonomous vehicles have been avidly studied in the past decade, as the trend tends toward fully independently operating systems. This includes a multitude of mobile, heavy-duty machines that utilize large-scale manipulators (also called booms) to complete work tasks. In underground mining applications, such manipulators are utilized in tunneling jumbos and drill rigs, for example. Although these machines are currently human operated, on-site or remotely via teleoperation, each joint has a sensor so that the tool center point (TCP) of the manipulator can be measured based on the joint states. This is because knowledge of the TCP pose is essential for accurate drilling, as it has a direct effect on the mining progress made with each blast. The number of actuator degrees-of-freedom (DOF) in these manipulators is typically up to 8, which means that 8 sensors, including their waterproof enclosures, associated cabling, and high-precision mechanical couplings, add up to a significant bill of materials, and thus, the cost per machine. Moreover, a single machine can have multiple

booms. For instance, tunneling jumbos typically have one to three drilling booms, which further add to the bill of materials and motivate research for new, alternative TCP measurement methods designed for this type of application.

Such methods have been explored for many similar applications in the literature. For example, in [1], a laser scanner was used for end-effector tracking and joint variable extraction of a heavy mining shovel's dipper. In [2], a low-cost 2D laser scanner was used to estimate the posture of a forestry crane, with a reported average tip position accuracy of 4.3 cm. With scanner-based measurements, the sparsity of the acquired point cloud becomes an issue at larger distances due to the low spatial resolution of the sensor. A part detection-based scheme for estimating the 2D pose of an excavator was studied in [3]. The method used a database of synthetic images comprising different parts of an excavator, which were used to train part detectors. A single camera was then used for extracting the skeleton of the excavator based on the detected parts. The feasibility of a local positioning system for loader cranes using wireless sensors was studied in [4]. The described method used inertial sensors for joint angle measurements and an ultrasonic transducer for measuring the length of a telescopic joint. In [5], a gravity-referenced joint angle estimation scheme using three-axis linear accelerometers and three-axis rate gyros was proposed, with reported joint angle sensing errors of ± 1 degree. An ultra-wide band (UWB) based real-time location system for estimating crane poses was studied in [6]. However, based on the UWB system error alone, which was approximately 30 cm, the accuracy is not sufficient for applications requiring precise positioning. In [7], an optical marker-based end-effector pose estimation scheme was presented for articulated excavators, with encouraging initial results. Nonetheless, these sensing methods do not particularly fit the present application of interest, as drilling booms are typically complex structures (8 actuator DOF) with considerable maximum lengths. The confined workspace also restricts the placement of sensors, such as cameras, in the environment around the machine.

A potential solution would be to place a camera directly at the TCP and perform pose estimation based on visual odometry or more advanced simultaneous localization and mapping (SLAM) algorithms. SLAM has attracted considerable attention in the past few decades, as this technology is a vital component of any autonomous vehicle: A machine cannot operate independently unless it is aware of its location in relation to the environment. Thus, the main objective of visual SLAM is to constantly perform localization based on visual feeds, while simultaneously building a map of the surroundings. Only recently have SLAM technologies showed signs of advancing toward the levels of maturity and reliability that are required for autonomous systems. Some areas, however, such as fail-safe systems, are still relatively unexplored, as discussed in a recent survey paper [8].

Numerous SLAM methods have been engineered over the years, with previous ones surveyed in [9]. Some of the more recent and most popular monocular schemes include DVO [10] and SVO [11]. Stereo and RGB-D methods have also been presented, for example, RTAB-MAP [12], in which the authors also provided comparative results using many SLAM algorithms available on ROS. A stereo SLAM method using ORB [13] features and line segments was proposed in [14]. The rationale was that the inclusion of line segments improves the performance in low-textured environments with planar structures, where a low number of point features can be extracted. The reported performance was similar to that of ORB-SLAM2 [15], which is another SLAM method. ORB-SLAM2 is fully based on ORB features, and can be used with mono, stereo, or RGB-D input. The authors extended their work to include inertial sensors in [16].

In this paper, inspired by the recent advances in SLAM technologies in vehicle positioning, we study the application and feasibility of SLAM in a much different setting. Specifically, we apply SLAM to estimate the TCP pose of a large-scale articulated crane in an unknown, confined space. The motivation is that for mining manipulators underground, the workspace area is typically small and confined, with walls closing in on each direction. The laboratory-grade simulation of such an environment is a test wall built from decorative stones. We attach a low-cost stereo camera near the TCP of an articulated heavy-duty crane so that the camera faces the test wall. Based on the literature review and the availability of state-of-the-art open source SLAM algorithms, we utilize the feature-based ORB-SLAM2¹, which is a tried and tested algorithm with excellent localization capabilities in varying environments. To avoid the issue of scale ambiguity, stereo vision is employed in the experiments. Data analysis is performed by comparing ground-truth TCP poses, obtained using joint encoders, with calibrated SLAM output poses. For offline data analysis, the calibration between the SLAM poses and the ground-truth poses is conducted using an iterative closest point (ICP) algorithm.

The remainder of this paper is outlined as follows: In Section II, a description of the application is provided; it is followed by Section III, in which the experimental setup is presented. In Section IV, the data analysis is presented with

comparative results. Finally, in Section V, the conclusion is provided.

II. SIMULTANEOUS LOCALIZATION AND MAPPING IN THE PROPOSED APPLICATION

Development drilling, in which tunnel networks are formed, is one of the basic operations in underground mining. Using this as an example, the TCP of a drilling boom is typically moved within an area resembling a rectangle that corresponds to the profile of the tunnel being mined. Dozens of holes, many meters in depth, are drilled inside the profile. The TCP is driven from drilling point to drilling point in a pre-planned manner, while during a drilling operation the current TCP pose is maintained. As the TCP is moved within a small area, with back-and-forth motions, it is possible to establish a comprehensive local map with SLAM before even beginning the drilling. This also suggests that loop detection and closing features of SLAM can be highly useful in correcting any drift in the pose estimates. Especially for the proposed application of TCP pose estimation, it is desirable that the tracking can be maintained at all times. Notably, heavy drilling induces severe vibrations in the manipulator and the machine, which could affect the tracking performance. In this case, having a local map would be useful, as relocalization based on the map can be instantly performed after a drilling operation, during which the joint positions can be locked in place by the control system. After the drilling plan is completed, explosive charges are placed inside the drill holes, after which blasting occurs. As the depth of a drill hole is measured in meters, any error in the TCP's pose will directly degrade the blasting result. Regarding the desired accuracy in this type of application, the rough target values for positioning and orienting are 1 cm and 1°, respectively.

The SLAM we utilize in this work, ORB-SLAM2, is a feature-based method that utilizes only ORB features, which have quickly become a popular choice due to their computational efficiency and good invariance to viewpoint changes. ORB-SLAM2 consists of three parallel threads: tracking, local mapping, and loop closing. If the tracking is lost, the system is capable of relocalization using a bag-of-words place recognition module, which is based on DBoW2 [17]. A pure localization mode is also available, in which the mapping and loop closing features are disabled. Importantly, the system applies bundle adjustment (BA) for optimization purposes at various stages of the algorithm: for optimizing the camera's orientation and position by minimizing the reprojection error between matched 3D landmarks and 2D key points in the tracking thread (motion-only BA); for optimizing a window of keyframes and points in the local mapping thread (local BA); and after a loop closure for optimizing all keyframes and points (full BA) [15]. The Levenberg-Marquardt method used for optimization is implemented in [18].

III. EXPERIMENTAL SETUP

A. System Description

The manipulator we used for testing was a HIAB033, which is a hydraulically actuated crane. The manipulator, illustrated in Fig. 1, additionally had a spherical wrist with a gripper attached to it, yielding a total of 6 active actuator DOF: rotate,

¹https://github.com/raulmur/ORB_SLAM2

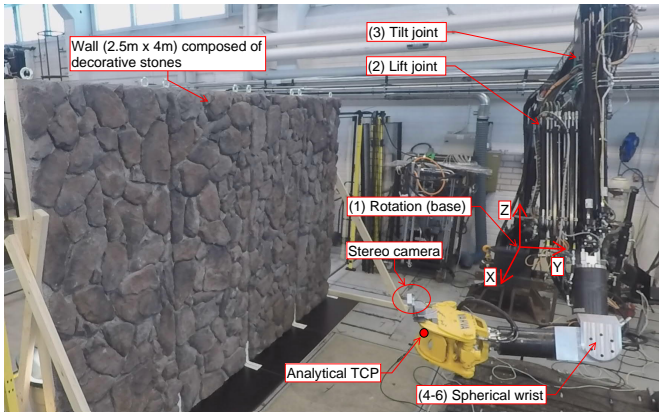


Fig. 1. The figure illustrates the test setup, in which the manipulator was positioned so that the stereo camera attached to the gripper faced the test wall. The goal was to estimate the TCP pose of the manipulator based on pure visual SLAM, which extracted the required features from the test wall. The joints of the manipulator are labeled from 1 to 6, with the base coordinate system also shown.

lift, tilt, and 3 DOF in the wrist. The manipulator also had two extension cylinders, which were disabled during the tests. Each active joint was instrumented with an incremental encoder.

The control system of the HIAB was a dSPACE DS1005 PPC controller board and a development PC, which ran in 3 ms sampling time. A Stereolabs ZED stereo camera was used for visual measurements. The ZED was connected to a laptop via its USB interface, and images were captured using ZED SDK's Matlab plug-in by using a UDP trigger signal, which was established from the dSPACE development PC to the ZED laptop. The UDP trigger signal was transmitted at 8×3 ms time intervals, which was dictated by the time it took for the ZED SDK to capture and save a pair of grayscale, 672×376 resolution images. The trigger signal ensured that the image data recorded with the laptop and the encoder data recorded with dSPACE could be synchronized with each other.

For the SLAM experiments, a wall made out of decorative stones was built. This was our laboratory-grade simulation of a mine wall. The wall was 2.5×4 m in dimensions, and the stones were cemented to the wall randomly, albeit they comprised some recurring shapes. Varying motions were applied to the manipulator, and data was recorded using the camera and the joint encoders. Then, the corresponding image sequences were extracted from the data. Each sequence library was then processed by the ORB-SLAM2 stereo algorithm, which provided a pose sequence corresponding to each input sequence. In the SLAM settings, the number of ORB features was set to 2000, and the camera FPS was set to 41.6667, according to the UDP trigger signal. Fig. 2 shows the ZED stereo camera attached to the gripper of the manipulator, as well as an example view of the detected ORB features from the test wall.

B. Ground-truth TCP Pose

As a ground truth, or reference, measurement of the TCP pose is required to evaluate the estimated SLAM poses, a kinematic model of the manipulator was formulated. The states of the six active joints, measured with encoders, were then used with the forward kinematic model of the manipulator to

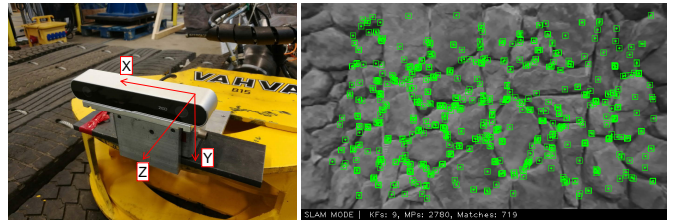


Fig. 2. The left image shows the stereo camera attached to the gripper. The camera's coordinate system is also shown. The right image displays an example view of the detected ORB features during SLAM.

TABLE I
DH PARAMETERS USED FOR TCP POSE FORMULATION.

No.	Joint	α_i	a_i	θ_i	d_i
1.	Rotation	$\pi/2$	a_1	θ_1	d_1
2.	Lift	0	a_2	θ_2	0
3.	Tilt	$\pi/2$	a_1	$\theta_3 + \pi/2$	d_3
4.	Wrist 1	$\pi/2$	0	θ_4	d_4
5.	Wrist 2	$-\pi/2$	0	θ_5	0
6.	Wrist 3	0	0	θ_6	d_6

formulate the TCP pose, which was used as the ground-truth measurement.

Table I presents the Denavit-Hartenberg (DH) parameters of the manipulator, which comprised an anthropomorphic arm with a spherical wrist [19]. The exact parameters were not used, as they were not available from the manufacturer. Instead, the parameters were self-measured, and are presented only symbolically here. The forward kinematic relationship between the base and the analytical TCP of the manipulator (see Fig. 1) was then formulated as follows:

$${}^1\mathbf{T}_6 = \mathbf{T}_1\mathbf{T}_2\mathbf{T}_3\mathbf{T}_4\mathbf{T}_5\mathbf{T}_6, \quad (1)$$

where ${}^1\mathbf{T}_6$ denotes the transformation matrix between the base frame and the TCP frame, and \mathbf{T}_i , $i \in \{1, \dots, 6\}$, was formulated using the following general equation by substituting the DH parameters of the i th joint:

$$\mathbf{T}_i = \begin{bmatrix} c\theta_i & -s\theta_i c\alpha_i & s\theta_i s\alpha_i & a_i c\theta_i \\ s\theta_i & c\theta_i c\alpha_i & -c\theta_i s\alpha_i & a_i s\theta_i \\ 0 & s\alpha_i & c\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

where \cos and \sin are abbreviated as c and s , respectively. The ground-truth TCP pose was then extracted from the ${}^1\mathbf{T}_6$ transformation matrix.

IV. DATA ANALYSIS

A. Extrinsic Calibration

As visual SLAM estimates the pose with respect to the camera frame, calibration between the camera frame and the ground-truth TCP frame is required to obtain comparable results. In more detail, the transformation matrix between the camera's coordinate system (or frame) and the coordinate system of the ground-truth TCP must be known, as the two are not inherently aligned. This results from the ambiguous camera attachment and the camera model. For the initial experiments presented in this paper, we used only recorded data, which

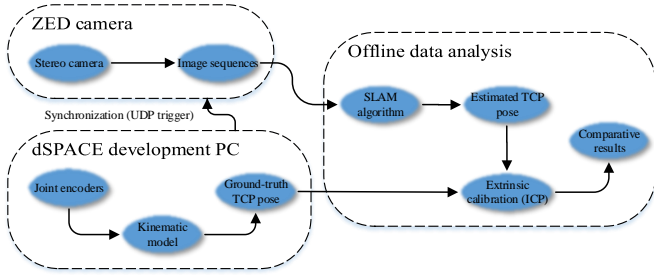


Fig. 3. A diagram illustrating how the comparative results were obtained from the measurements.

permitted the use of the ICP algorithm [20]. In essence, the SLAM pose trajectory is modified into a point cloud, which is then matched to the respective ground-truth pose trajectory point cloud by using an ICP search.

However, a weakness of the ICP is that it can provide an erroneous fitting, while the mean error between the two matched point clouds appears small. In attempt to avoid such a scenario, the camera frame was first modified so that the positive direction of each axis corresponded to that of the ground-truth TCP frame. The second step was to use the ICP to find the transformation matrix between the two frames.

B. Comparative Results

The performance of SLAM in estimating the TCP pose was experimented with motions into different directions. The TCP poses are expressed with respect to the base coordinate system (see Fig. 1), and the manipulator was automatically driven to the same initial TCP position before each experiment. The initial poses between experiments had minor variances, because P-control was used when the joints were driven. The procedure for how the results were obtained is further visualized in Fig. 3.

The first three experiments observed the orientation of the TCP with respect to each axis, whereas in the fourth and final experiment a longer trajectory with loops and multiple laps was studied. In the results, black lines always represent ground-truth variables obtained using encoder measurements and forward kinematics, whereas red lines represent calibrated SLAM estimates in each case. Furthermore, orientation is expressed with XYZ Euler angles.

The camera frame was first aligned with the ground-truth TCP frame using the ICP procedure described in the previous subsection. Static biases with respect to the ground-truth initial poses were also removed from the estimates. Table II shows the root mean square errors calculated during the ICP procedures. The errors are very small, implying that the calibrated camera frame should closely match the ground-truth TCP frame so that the poses are comparable.

In the first experiment, motion was applied only to the lift joint (see Fig. 1) so that the TCP rotated about the Y-axis. The resulting 6 DOF TCP pose is shown in Fig. 4, in which the translational motions and the respective orientations in relation to each axis are illustrated. The positional variables and the Y-axis orientation demonstrate good matching with their respective ground-truth measurements. The remaining two orientation estimates from SLAM show larger amplitudes of motion than their corresponding ground-truth measurements.

In the second experiment, motion was applied only to the base rotation. In this case, the camera (and the TCP) moved mainly in the depth direction along the Y-axis and around the Z-axis of the base frame. The outcome is illustrated in Fig. 5. The results are similar to those of the first experiment: The positional variables and the orientation of the main motion axis show good matching with the ground-truth measurements, while the other two orientations display larger motions.

In the third experiment, the goal was to rotate the TCP around the X-axis by moving the second wrist joint. With the present test setup, however, achieving rotational motion purely around the X-axis was not possible due to the wrist's structure. The TCP was also lifted upward before the second wrist joint was moved, which was to allow larger motion while the wall is maintained in the camera's view. The results are illustrated in Fig. 6. As shown, the positional variables and the main orientation match well in this case also. However, there are slight differences in the remaining two orientation variables, with the estimated angles displaying larger amplitudes.

For the fourth experiment, a TCP trajectory with multiple loops and laps was designed so that the loop detection and closing features of the SLAM algorithm could be tested. The complete path is illustrated in Fig. 7, where the black point cloud represents the ground-truth TCP trajectory. It is compared with the red point cloud, which visualizes the calibrated TCP trajectory obtained from SLAM. The first rectangular part of the path in the XZ plane was completed 3 times, after which the TCP was moved closer to the wall along the Y-axis. Then, the second rectangular part of the path in the XZ plane was also completed three times. Finally, the TCP was driven back to the initial position along the Y-axis. As the results in Fig. 7 show, the multiple laps during each loop in the XZ plane are barely visible in the point clouds and the loops are also closed. The respective 6 DOF TCP pose during the measurement is illustrated in Fig. 8. The results are in line with the previous experiments; the positional variables match well with the corresponding ground-truth measurements. The orientations also match relatively well, albeit the estimated angles show larger amplitudes of motion by a few degrees in relation to their ground-truth measurement counterparts.

To sum up the results of the four experiments, ORB-SLAM2 performed surprisingly well in the tested cases. The mean absolute errors of the TCP pose variables in each experiment (1–4) are documented in Table III, where $\gamma_{x,y,z}$ denote the XYZ Euler angles. Respectively, the maximum absolute errors are shown for each case in Table IV. The estimated orientation angles generally demonstrated larger amplitudes than the ground-truth measurements, which is expected to be at least partly attributed to the flexibility of the manipulator. The differences could also have followed from inaccuracies in the DH parameters (namely, the angles) or in the calibration step.

Finally, this work considered only a laboratory setting. In addition, a specifically designed test wall with a relatively textured surface was used. As visual SLAM is completely dependent on what the camera has in its view, the performance is strictly tied to the environment. Thus, real-world measurements from actual mines are required for further experimenting. In this work, we also used a stereo camera to avoid scale ambiguity, which is a well-known issue with

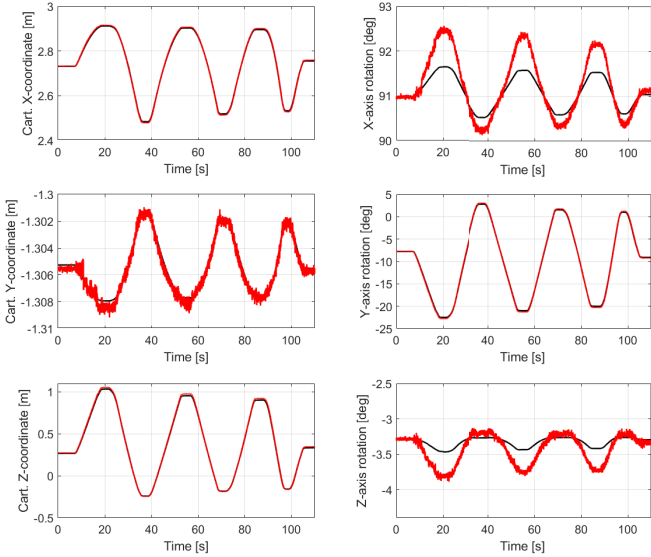


Fig. 4. Results from the first experiment, in which motion was applied only to the lift joint. The black lines denote the ground-truth values obtained with encoders, while the red lines denote the calibrated SLAM estimates.

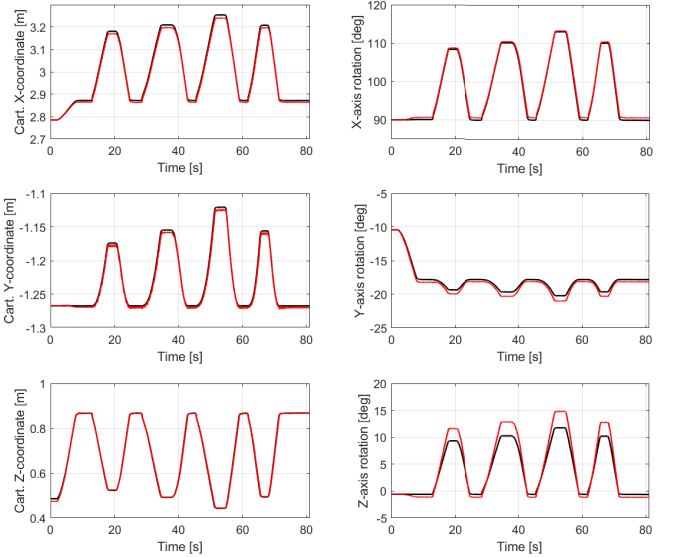


Fig. 6. Results from the third experiment, in which motion was applied mainly to the second wrist joint.

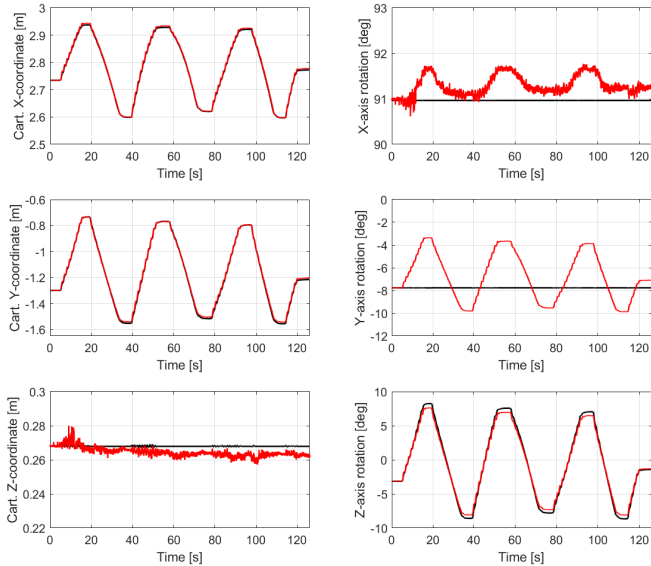


Fig. 5. Results from the second experiment, in which motion was applied only to the base rotation.

TABLE II
ICP ROOT MEAN SQUARE ERROR IN EACH CASE.

Experiment	1	2	3	4
RMS error	0.0032 (m)	0.0081 (m)	0.0038 (m)	0.0190 (m)

monocular systems. However, stereo implies that the system has a minimum viewing distance required for reliable triangulation of the 3D point features, which is not optimal for confined spaces. A possible solution would be to switch to monocular SLAM when the minimum distance is crossed, while the scale is obtained using stereo data or another sensor.

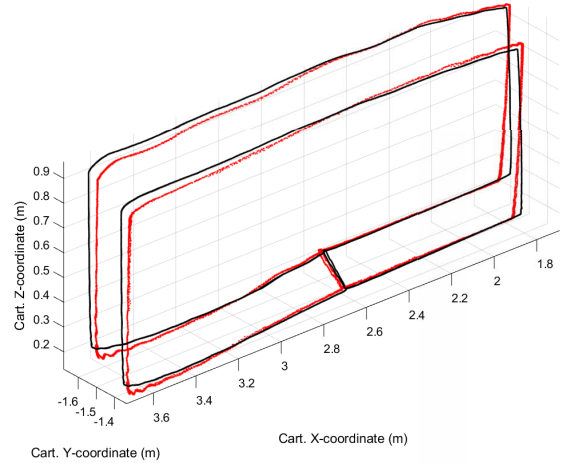


Fig. 7. In the fourth experiment, a longer TCP trajectory was experimented with. The first rectangular part of the path in the XZ plane was completed three times, after which the TCP was driven closer to the wall along the Y-axis. Then, the second rectangular part of the path in the XZ plane was also completed three times. Finally, the TCP was driven to the initial pose along the Y-axis. The black point cloud illustrates the ground-truth poses obtained with encoders, while the red point cloud illustrates the calibrated SLAM poses.

TABLE III
MEAN ABSOLUTE ERRORS IN EACH MEASUREMENT.

	Fig. 4	Fig. 5	Fig. 6	Fig. 8
x (m)	0.0032	0.0033	0.0082	0.0338
y (m)	0.0003	0.0073	0.0027	0.0056
z (m)	0.0089	0.0038	0.0032	0.0133
γ_x (deg)	0.2746	0.3352	0.4306	0.8988
γ_y (deg)	0.1565	1.8851	0.4040	0.7108
γ_z (deg)	0.1312	0.3589	1.0950	0.4824

V. CONCLUSION

In this work, we studied the application and feasibility of SLAM for estimating the TCP pose of a large-scale manipulator in a confined, unknown environment. The SLAM

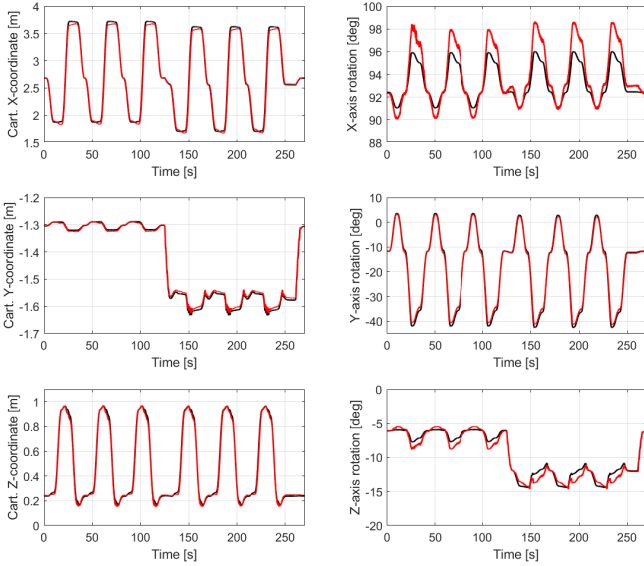


Fig. 8. Results from the fourth experiment.

TABLE IV
MAXIMUM ABSOLUTE ERRORS IN EACH MEASUREMENT.

	Fig. 4	Fig. 5	Fig. 6	Fig. 8
x (m)	0.077	0.0098	0.0203	0.1317
y (m)	0.0015	0.0246	0.0097	0.0229
z (m)	0.0239	0.0116	0.0129	0.0492
γ_x (deg)	0.9166	0.8064	0.9014	2.6606
γ_y (deg)	0.3451	4.4364	0.8373	2.1485
γ_z (deg)	0.4268	0.9145	3.0922	1.4357

algorithm is a key part of the proposed application, in which accuracy, robustness, and real-time performance are all highly important. In the initial results presented in this paper, we were mainly concerned about the potential accuracy. We found that ORB-SLAM2 provided a relatively good performance in the offline data analyses, in which we used an ICP algorithm to extrinsically calibrate the camera. Based on previous research, ORB-SLAM2 is also directly applicable in real time.

Regarding the calibration procedure, the results benefit from running an automatic calibration algorithm by applying ICP to the outputs of each individual experiment. This allowed the effective fine-tuning of the extrinsic calibration parameters of the camera for each test by comparison with the ground truth. In the case of online tests and use of SLAM feedback for control purposes without the ground truth, further investigation of the system calibration is required for online estimation of the extrinsic parameters.

Although the results were obtained specifically with ORB-SLAM2, theoretically the SLAM algorithm itself should not matter as long as the 6 DOF TCP pose can be reliably estimated. To fully localize the TCP of a manipulator with respect to the machine it is attached to, SLAM by itself is not sufficient, as it estimates only the motion of the camera (or the TCP) frame. In this work, the relationship between the frame and the base frame of the manipulator was obtained from ground-truth joint encoder measurements. For future studies,

the goal is to omit these sensors completely by developing an alternative method for formulating this correspondence.

REFERENCES

- [1] A. H. Kashani, W. S. Owen, N. Himmelman, P. D. Lawrence, and R. A. Hall, "Laser scanner-based end-effector tracking and joint variable extraction for heavy machinery," *The International Journal of Robotics Research*, vol. 29, no. 10, pp. 1338–1352, 2010.
- [2] H. Hytti, V. V. Lehtola, and A. Visala, "Forestry crane posture estimation with a two-dimensional laser scanner," *Journal of Field Robotics*, vol. 35, no. 7, pp. 1025–1049, 2018.
- [3] 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.
- [4] P. Cheng, B. Oelmann, and F. Linnarsson, "A local positioning system for loader cranes based on wireless sensors—a feasibility study," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2881–2893, 2011.
- [5] J. Vihonen, J. Mattila, and A. Visa, "Joint-space kinematic model for gravity-referenced joint angle estimation of heavy-duty manipulators," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 12, pp. 3280–3288, 2017.
- [6] 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, 2011.
- [7] K. M. Lundeen, S. Dong, N. Fredricks, M. Akula, J. Seo, and V. R. Kamat, "Optical marker-based end effector pose estimation for articulated excavators," *Automation in Construction*, vol. 65, pp. 51–64, 2016.
- [8] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [9] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 55–81, 2015.
- [10] C. Kerl, J. Sturm, and D. Cremers, "Dense visual SLAM for RGB-D cameras," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 2100–2106.
- [11] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 15–22.
- [12] M. Labbé and F. Michaud, "RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.
- [13] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *ICCV*, vol. 11, no. 1. Citeseer, 2011, p. 2.
- [14] R. Gomez-Ojeda, F.-A. Moreno, D. Zuñiga-Noël, D. Scaramuzza, and J. Gonzalez-Jimenez, "PI-slam: a stereo slam system through the combination of points and line segments," *IEEE Transactions on Robotics*, 2019.
- [15] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [16] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular slam with map reuse," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, 2017.
- [17] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, October 2012.
- [18] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3607–3613.
- [19] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: modelling, planning and control*. Springer Science & Business Media, 2010.
- [20] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures*, vol. 1611. International Society for Optics and Photonics, 1992, pp. 586–607.