Redundancy-Based Visual Tool Center Point Pose Estimation for Long-Reach Manipulators

Petri Mäkinen, Pauli Mustalahti, Sirpa Launis and Jouni Mattila

Abstract— **In this paper, we study a visual sensing scheme for 6 degree-of-freedom (DOF) tool center point (TCP) pose estimation of large-scale, long-reach manipulators. A sensor system is proposed, designed especially for mining manipulators, comprising a stereo camera running a simultaneous localization and mapping (SLAM) algorithm near the TCP and multiple cameras that track a fiducial marker attached near the stereo camera. In essence, the TCP pose is formulated using two different routes in a co-operative (eye-in-hand/eye-to-hand) manner using data fusion, with the goal of increasing the system's fault tolerance and robustness via sensor redundancy. The system is studied in offline data analysis based on real-world measurements recorded using a hydraulic 6 DOF robotic manipulator with a 5 m reach. The SLAM pose trajectory is obtained using the open source ORB-SLAM2 Stereo algorithm, whereas marker-based tracking is realized with a high-end motion capture system. For reference measurements, the pose trajectory is also formulated using joint encoders and a kinematic model of the manipulator. Results of the 6 DOF pose estimation using the proposed sensor system are presented, with future work and key challenges also highlighted.**

I. Introduction

A. Motivation

Mobile working machines represent a significant field in industry, and they come in many different configurations and sizes with respect to their on-board manipulators. In machines designed specifically for mining and construction, the reach of these manipulators can range from approximately 10–15 m in 6 degrees-of-freedom (DOF) tunneling machines to only 1–2 m in small surface drilling platforms. The annual production volume for a specialized machine type can be a few hundred units, while the volume for some production variants can be as low as 1–10 units per year. Therefore, these high-precision, low-volume robotic manipulators call for innovative sensor system solutions that reduce the manufacturing, assembly, and maintenance costs of these machines. The current solution is to fit each joint of a manipulator with a joint sensor, which also requires additional protective housing, mechanical couplings, and cabling that are suitable for the given machine type. Therefore, many components are required to fit all the machine types with mechanical precision sensing, which also results in an overall high cost in the terms of the bill of materials (BOM). The underlying goal of the present research is that all types of mining machines are equipped with a standardized sensor system that is of low cost, easy to install, and scalable to fulfill all requirements across the range of machine types.

Petri Mäkinen, Pauli Mustalahti and Jouni Mattila are with the Faculty of Engineering and Natural Sciences, Tampere University, Tampere, Finland. Emails: {petri.makinen,pauli.mustalahti,jouni.mattila}@tuni.fi

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

Each manipulator should have a sensor system because the 6 DOF tool center point (TCP) pose of the manipulator must be known. In mining machines, knowledge of the joint states and the TCP pose is currently required to carry out automated and semi-automated operations, as a high production rate is very valuable. Due to this, tunneling jumbos, for example, can have up to four drilling booms attached to the same machine for parallel operations. Overall, the goal is to move toward fully automated operations in these hazardous applications, as discussed in, for example, [1]. From a practical point of view, the important factor is being able to accurately measure and control the TCP pose that is expressed in Cartesian work space with respect to a world frame. For this purpose, methods other than the traditional kinematic chain formulation with joint sensors could also be developed. Due to the long reach and high payload-to-own-weight ratios of these manipulators, the traditional method based on a serial kinematic chain structure will always impose significant errors at the end of the chain (the TCP) due to structural flexibilities and calibration uncertainties. Thus, driving these manipulators with external sensor systems, in GPS-denied environments, is of great interest.

Compared to traditional industrial robots, large-scale, long-reach manipulators are often under the radar in research. Whereas an industrial (stationary) robot has a relatively low payload-to-own-weight ratio and precision sensors at each joint providing the manufacturer-quaranteed absolute accuracy and repeatability for the TCP pose in Cartesian space, large-scale manipulators have much higher payload-to-own-weight ratios (e.g., one), with many applications still operated manually, as no sensors are installed due to the harsh working conditions and structural flexibilities that distort the results if basic rigid-body kinematics are applied. This situation is changing, however, as the automation level of these manipulators is increasing, thus requiring sensors. In forestry machines, for example, inertial sensors have been recently introduced commercially to measure the joint angles that are sensitive to gravity, making it possible for the operator to control the TCP directly, instead of controlling each individual joint of the manipulator. A method for computing gravity-sensitive angles using inertial sensors was introduced in [2].

In mining machines, sensors have long been present to measure the joint states, as in this application the TCP pose is required so that drilling plans can be effectively completed.

For example, an orientation error of 5° at the TCP, with a drilling depth of 4 m, will result in a position error of 35 cm at the end of a drill hole. The accuracy of the drill holes with respect to the drilling plan is crucial. In tunneling, inaccurate drilling results in more drilled meters required, along with more blastings required, which slows progress and increases operation costs. Respectively, in long-hole drilling, inaccurate drilling can lead to ore-loss or increased dilution (waste rock) of the product. Overall, straight drill holes result in a better total economy. As rough target values for accuracy in these applications, the positioning error at the TCP should be less than 1 cm and the orientation error less than 1°, respectively.

This paper is the new step after our previous research [3], in which pure visual simultaneous localization and mapping (SLAM)-based TCP pose estimation was studied. In this paper, we extend toward a more complete sensor system concept for the described application. Namely, marker-based TCP pose tracking is combined with the SLAM module in an attempt to obtain a more robust pose estimation. In essence, this corresponds to the so-called eye-in-hand/eye-to-hand co-operation, which is a method used for visual servoing, see e.g. [4], [5]. In this work, the goal of the proposed solution is to increase the system's fault tolerance in the sense of sensor redundancy, while having both measurements (marker tracking and SLAM) available complement each other after data fusion. For marker tracking, we used a commercial OptiTrack motion capture system, which conveniently offered the required functionality for measurements, such as calibration and multi-camera tracking. Although such a high-end system in commercial mining machines is unrealistic, lower-end cameras are becoming more affordable and advanced, not only the hardware but also software. Consequently, multi-camera solutions with redundancy are becoming more viable in cost-sensitive industry applications, and this paper is a step toward this path. For SLAM, we used a Stereolabs ZED camera along with the open source ORB-SLAM2[1] Stereo algorithm [6]. A test case using a laboratory-installed 6 DOF hydraulic crane was designed for a simple practical experiment to study the feasibility and challenges in realizing the conceptual sensor system at full scale. The results of the offline data analysis show that the main challenges lie in the system's calibration (for precise measurements) and in control design. Further practical issues, such as model development for kinematic calibration of flexible robots, are beyond the scope of this study.

B. Brief Literature Review

Due to the harsh and highly varying environmental conditions that large-scale mobile manipulators are exposed to, a wide variety of sensor technologies have been explored. For example, in [7] battery-powered wireless sensors were applied for local positioning of a loader crane. Inertial sensors were used to measure joint angles, and an ultrasound time-of-flight sensor was used to measure the length of a telescopic extension boom. In [8], a laser scanner was used with a customized iterative closest point algorithm to estimate the joint angles.

A marker-based pose estimation method for articulated excavators was presented in [9]. In this case, the camera was installed in the surrounding environment, and several challenges were brought up, such as occlusion and lighting. Thus, a marker-based system is foreseen to work best as an auxiliary sensing method. A marker-less method for the same problem was later presented in [10], in which a deep convolutional network human pose algorithm was used. These studies have in common that they focused on articulated manipulators that have joints only in the vertical plane. Mining manipulators, however, typically also have at least two joints in the horizontal plane, which complicates the pose estimation problem significantly.

The rest of the paper is organized as follows: Section II describes the proposed sensor system, whereas Section III details the experimental setup for measurements. It is followed by Section IV, which contains data analysis and results. Finally, Section V concludes the paper.

II. Conceptual Sensor System for Mining Manipulators

Inertial sensors cannot be effectively utilized in mining manipulators due to the presence of several horizontal joints that are insensitive to gravity. As for visual sensing, such systems are already utilized for collision avoidance [11]. The sensors are installed near the roof of the cabin of a machine. Moreover, unlike with articulated cranes found in, for example, excavators, in mining manipulators the base of the manipulator is typically lower with respect to the cabin, as shown in Fig. 1, giving natural elevation to visual sensors installed near the roof. This results in fewer occlusions that would result due to a part of the manipulator blocking the TCP. However, as discussed in [9], pure marker-based systems can be problematic to realize due to several reasons, which suggests that additional sensors are required for increasing accuracy and robustness.

In the previous study [3] we used SLAM to estimate the TCP pose with good initial results. The idea was that the surrounding mine walls provide enough features, and that the operations conducted by these mining and construction machines are controlled enough for SLAM to be viable. It is also perceived that due to the length of these manipulators ($>$ 10 m), a sensor located near the TCP is required to obtain precise measurements, which is supported by the results of this paper. Thus, a sensor system resembling the eye-in-hand/eye-to-hand co-operative scheme is studied in this work. The SLAM module, located near the TCP of a manipulator, can provide more precision in pose estimation. The marker-based tracking module will be less accurate due to the increased viewing distances, and thus, increased uncertainties, but it will have a better view of the entire scene, which could also be used for calibration and other assistive operations, for example.

[1] https://github.com/raulmur/ORB_SLAM2

Fig. 1. A Sandvik tunneling jumbo with two drilling booms.

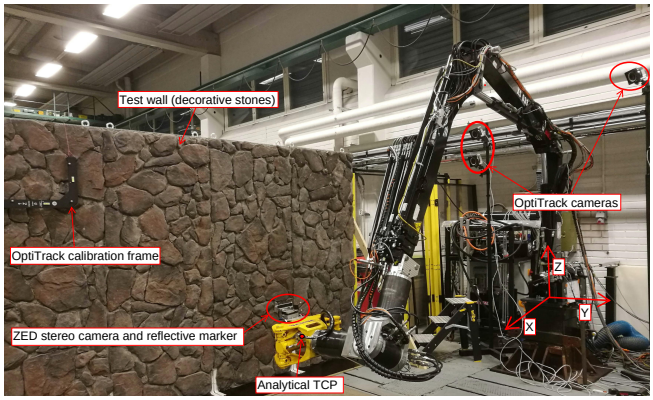

Fig. 2. The experimental measurement setup. The goal was to track the TCP pose of the manipulator by using i) SLAM (ZED stereo camera), in which the tracked features were obtained from the test wall, and ii) marker-based tracking, for which a high-end OptiTrack motion capture system was used. For reference measurements, an analytical TCP of the manipulator was also formulated based on forward kinematics and joint encoder measurements. The world coordinate system is also shown.

III. Experimental Setup

To study an application with the proposed sensor system architecture and obtain initial results, a simple use case test bed was designed. The system comprised the following main components:

- A laboratory-installed hydraulic crane with accurate reference sensors and a dSPACE real-time control system.
- A Stereolabs ZED camera for SLAM, along with a textured test wall for feature extraction.
- An OptiTrack motion capture system for marker-based tracking.

The components are detailed further in the next subsections.

A. HIAB033 Hydraulic Crane with Additional 3 DOF Wrist

The target system was a hydraulic lorry crane, HIAB 033, which was located at the heavy laboratory of the Innovative Hydraulics and Automation research unit at Tampere University. The setup is presented in Fig. 2. The manipulator itself had 3 active DOF (rotation, lifting, and tilting). A spherical wrist was also attached at the tip of the structure, adding another 3 DOF to the system. Each of the six active joints was instrumented with an incremental encoder to obtain precision measurements of the joint states.

B. SLAM Module

To estimate the TCP pose with SLAM, a Stereolabs ZED camera was installed near the tip of the manipulator. Grayscale images were captured at 24 ms intervals with a resolution of 672×376 per lens and saved for offline data analysis. As for the SLAM method, we utilized the open source ORB-SLAM2 Stereo algorithm. For the textured environment, from which the feature points for SLAM were to be obtained, a 2.5×4 m test wall was constructed using decorative stones. The underlying goal was to simulate a rock wall, as the target application of this research was underground mining and construction.

C. Marker-Based Tracking Module

As this study was mainly concerned about a conceptual sensor system, we used the most powerful systems available to us. In this case, we used a commercial-off-the-shelf motion capture system to realize high-performance marker tracking. Three OptiTrack Prime 17W cameras were placed around the base pillar of the manipulator: The idea is that the cameras used for marker-based tracking are installed on top of the cabin of a machine. An infrared-reflective (passive) marker was then placed near the tip of the manipulator (next to the ZED camera). Using OptiTrack's Motive motion capture software, the marker's pose was tracked with reference to an OptiTrack L-frame, which was placed next to the test wall and in view of the cameras. With three cameras, the system was able to effectively track the marker, although the boom temporarily occluded the view of the third camera during the measurements.

D. Data Flow

A general depiction of the data flow during the measurements is presented in Fig. 3. The real-time control system of the manipulator was a dSPACE DS1005 PPC controller board, which used a 2 ms sampling period and recorded the encoder and motion capture measurements. The OptiTrack cameras were read using a dedicated laptop running Motive software, from which the measured poses were transmitted at a high frequency to the dSPACE development PC by using Matlab and UDP. The ZED stereo image capture was also realized with a dedicated laptop, which was synchronized with the dSPACE development PC by sending a UDP trigger signal from dSPACE to the dedicated laptop at a time interval of 12×2 ms. The synchronized image sequences were then recorded on the dedicated laptop by using Matlab and the ZED SDK.

E. Ground-Truth TCP Pose

To have a reference pose for the camera measurements, a kinematic model of the manipulator was formulated. This was then used with the encoder measurements to produce the TCP pose based on the model.

Remark 1: Compared to our previous study [3], the Denavit-Hartenberg (DH) parameters of the manipulator were kinematically calibrated using a Sokkia NET05 total

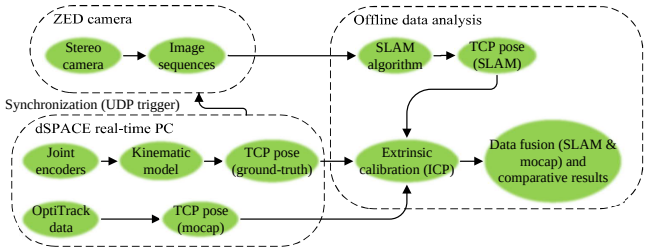

Fig. 3. A general depiction of how the data was obtained and used for offline data analysis.

TABLE I

DH PARAMETERS OF THE MANIPULATOR. THE SYSTEM COMPRISED AN ANTHROPOMORPHIC ARM IN COMBINATION WITH A SPHERICAL WRIST.

Joint	α_i	a_i	θ_i	d_i
Rotation	$\pi/2$	a_1	θ_1	d_1
Lift	0	a_2	θ_2	0
Tilt	$\pi/2$	a_1	$\theta_3 + \pi/2$	d_3
Wrist 1	$\pi/2$	0	θ_4	d_4
Wrist 2	$-\pi/2$	0	θ_5	0
Wrist 3	0	0	θ_6	d_6

station. The resulting Cartesian average error in the calibration was reportedly less than 4 cm.

The symbolic DH parameters are presented in Table I. The forward kinematic relationship between the base and the analytical TCP of the manipulator (see Fig. 2) was then formulated as follows:

$$ {}^{B}\mathbf{T}_{tcp} = \mathbf{T}_r\mathbf{T}_l\mathbf{T}_t\mathbf{T}_{w_1}\mathbf{T}_{w_2}\mathbf{T}_{w_3} \tag{1} $$

where the transformation matrix from the base to the TCP is denoted by ${}^{B}\mathbf{T}_{tcp}$. Joint transformation matrices $\mathbf{T}_i$, $i \in \{r, l, t, w_1, w_2, w_3\}$ are obtained using the following equation by substituting the respective DH parameters for each joint:

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

where $s = sin$ and $c = cos$. Then, the TCP pose of the manipulator is obtained from ${}^{B}\mathbf{T}_{tcp}$.

F. Data Fusion of the TCP Pose Estimates

The logic for obtaining the TCP pose estimates based on the available signals was designed as shown in Fig. 4. In the scope of this work, it is assumed that each sensor is either fully operational or not operational (binary), as self-diagnostic systems would be required for more advanced signal analysis. In the event that both TCP pose estimates are available, a data fusion method, confidence-weighted averaging [12], is adopted. This simple, model-free method fuses measurements based on the estimated variance of the measurement error. The advantage is that, assuming that the errors between the sensors are independent and that the

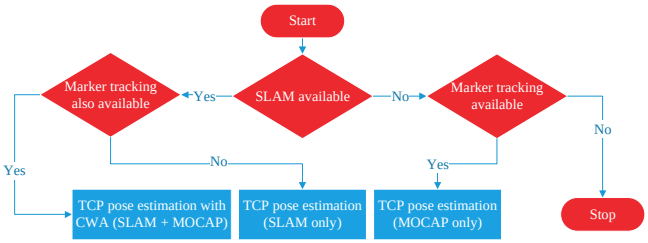

Fig. 4. A chart illustrating the logic behind utilizing the two TCP pose estimates. In the ideal case that SLAM and marker tracking (mocap) are available, confidence-weighted averaging (CWA) is used for data fusion.

expected error equals zero, the variance of the fused output is minimized. The fused 6 DOF TCP pose vector X_{fused} is obtained as follows:

$$ X_{fused} = \sum_{i=1}^{N} W_i X_i \tag{3} $$

where N denotes the total number of observations (in this case, SLAM and marker-based tracking), W_i is the weight vector of the ith observation, and X_i is the 6 DOF TCP pose vector of the ith observation. The weights are computed based on the signal variances as follows:

$$ W_i = \frac{1}{\sigma_i^2} \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \tag{4} $$

where $\sigma_{i,j}^2$ denotes the variance of a given signal.

Furthermore, if no TCP pose estimate is available, then the manipulator is halted. Matlab Simulink's Stateflow environment was utilized in the experiments.

IV. DATA ANALYSIS

A. Calibration of the TCP Frame Correspondences

To obtain comparable results, calibration between the three TCP frames (coordinate systems) is required. The SLAM frame and the marker frame are to be transformed into the analytical TCP frame, which served as the reference. For this purpose, the iterative closest point (ICP) method [13] was employed, which is suitable for offline experiments because the entire pose trajectories from different sources can be matched.

B. Signal Conditioning

The measured TCP orientations using marker-based tracking were conditioned with a geometric moving average (GMA) filter [14] due to noisy data. The equation is given as follows:

$$ S_j = (1 - \alpha)S_{j-1} + \alpha s_j, \quad j > 0 \tag{5} $$

where S_j denotes the geometric moving average (conditioned signal) at time j, s_j denotes the unconditioned signal at time j, and $0 < \alpha \leq 1$ denotes the weight coefficient. Note that S_0 is set to the initial value of a given signal. Furthermore, $\alpha = 0.02$ was used.

C. Comparison of Pose Trajectories

First, a test trajectory was designed for the manipulator by using quintic path planning [15]. A rectangular-shaped trajectory was completed three times, after which the TCP was moved closer to the test wall. Then, the rectangle was completed three times again. Finally, the TCP was moved back to the initial position. The trajectories are shown in Fig. 5, which illustrates the three TCP pose trajectories in Cartesian space after point cloud matching using the ICP. The black point cloud represents the analytical reference trajectory, the red point cloud represents the SLAM output poses, and the blue point cloud is associated with the trajectory of the tracked marker. The respective root mean square errors resulting from the ICP matching algorithm are presented in Table II.

The 6 DOF TCP pose estimates are presented based on the chart in Fig. 4. First, only the SLAM pose estimates were used, with the resulting 6 DOF poses shown in Fig. 6. Respectively, the 6 DOF poses from marker-based tracking are shown in Fig. 7. The CWA-fused 6 DOF pose estimates are shown in Fig. 8. The visual measurements in each case were compared with the reference encoder data, with the mean and maximum absolute errors documented in Table IV and Table V. Red lines are associated with the SLAM poses, blue lines denote the marker-based tracking, and black lines represent the encoder computed data. As shown by the measured results, the Cartesian position variables track relatively well over the entire test trajectory, with the mean errors ranging from less than a millimeter to a few centimeters. The orientation variables, however, show less consistent behavior as the amplitudes seize to match well after the TCP is driven closer to the wall. It is suspected this followed from calibration errors, as the uncertainties present in the reference encoder setup and the calibration were quite significant. Furthermore, the visual sensors provided similar behavior, with especially the OptiTrack system being perceived as capable of highly accurate measurements. This emphasizes the challenge of obtaining an accurate 6 DOF pose reference in large-scale, long-reach manipulators.

The two ICP calibrated optical measurements were also compared with each other, see, Fig. 9. Here, the unconditioned marker orientations are also shown with light-blue lines. The results demonstrate a strong correspondence between the SLAM poses and the marker poses. The mean and maximum absolute errors between the optical measurements were also documented in Tables IV-V.

It is evident that the visual estimates of the TCP pose differ from the analytical TCP based on the kinematic model, which suggests that transitioning from the optical measurements to the joint space of the manipulator will be challenging. Thus, alternative, external methods of controlling these manipulators, instead of using the numerical serial kinematic chain structure, should be explored. For the offline data analysis, the weights of the data fusion were obtained using the variances computed over the entire test trajectory by using the encoder reference measurements as the ground-truth values. The weights used are presented in Table III. As it shows, the marker-based tracking was emphasized in fusing the positions which, in this case, was logical due to the reduced ICP calibration error. The orientations were weighted quite evenly. However, the orientations from the marker-based tracking module were GMA-filtered before weights were computed. Consequently, the resulting data fusion is, in this case, optimal in the sense that the fused variance was minimized. However, for future online experiments, a method for determining the weights in real-time is required.

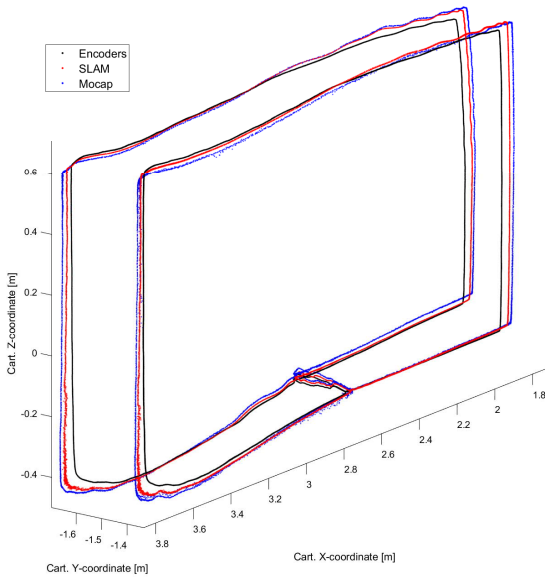

Fig. 5. The pose trajectories after ICP registration and calibration. A rectangular trajectory was first completed three times, after which the TCP moved closer to the test wall and completed another three laps on a rectangular trajectory. Finally, the TCP was moved back into the initial position. The black point cloud represents the analytical TCP, the red point cloud represents the SLAM TCP, and the blue point cloud represents the tracked marker TCP.

TABLE II
ROOT MEAN SQUARE ERRORS RESULTING FROM THE ICP ALGORITHM.

Coord. transf.	SLAM→Analytical TCP	Marker→Analytical TCP
RMS error	0.0371 [m]	0.0253 [m]

TABLE III
WEIGHTS USED IN THE CWA DATA FUSION.

W_i	x	y	z	γ_x	γ_y	γ_z
SLAM	0.2828	0.4728	0.1311	0.4767	0.6613	0.4948
Mocap	0.7172	0.5272	0.8689	0.5233	0.3387	0.5052

V. DISCUSSION AND CONCLUSION

This work presented a new sensor system concept designed especially for large-scale, long-reach mining and construction manipulators used underground, in which an eye-in-hand/eye-to-hand co-operative scheme is utilized by

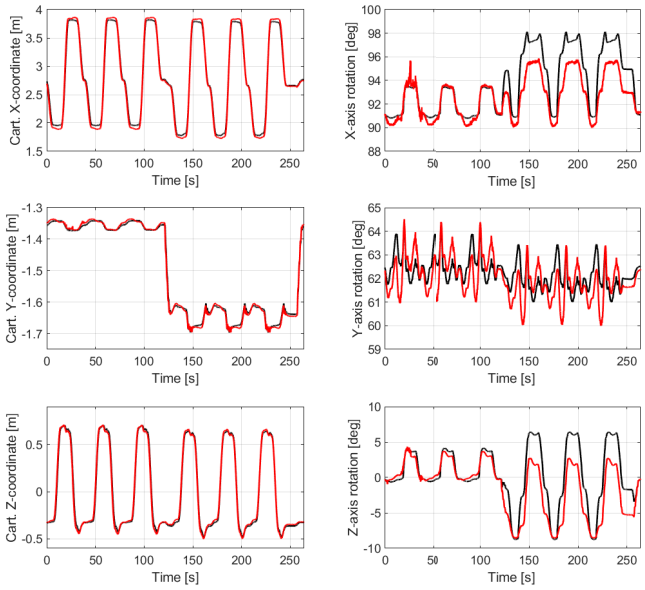

Fig. 6. The estimated 6 DOF TCP pose variables, when only SLAM is available. The black lines denote the reference values using encoder measurements, and the red lines denote the SLAM pose variables.

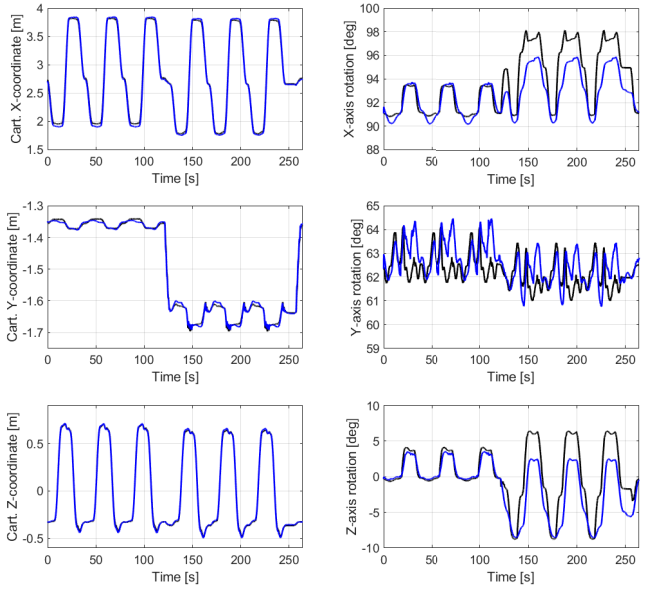

Fig. 7. The estimated 6 DOF TCP pose variables, when only marker-based tracking is available. The black lines denote the reference values using encoder measurements, and the blue lines denote the values from marker tracking, with GMA-filtered orientations.

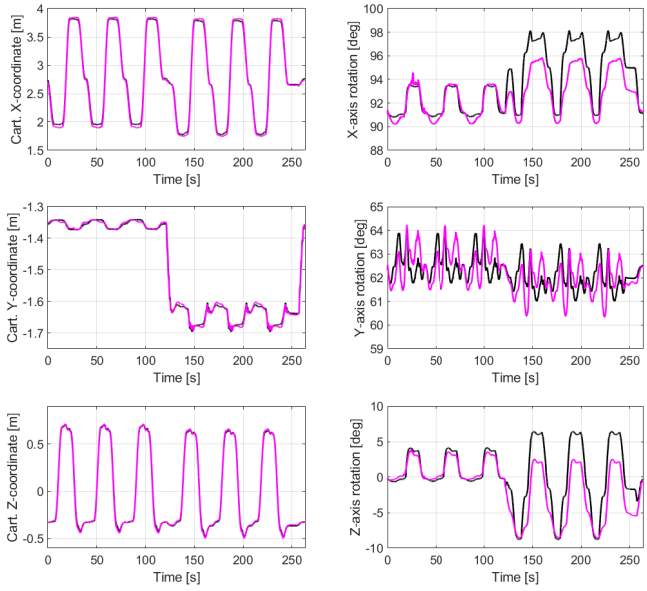

Fig. 8. The estimated 6 DOF TCP pose variables, with signals from SLAM and marker tracking fused with CWA. The black lines denote the reference values using encoder measurements, and the magenta lines denote the pose variables after data fusion.

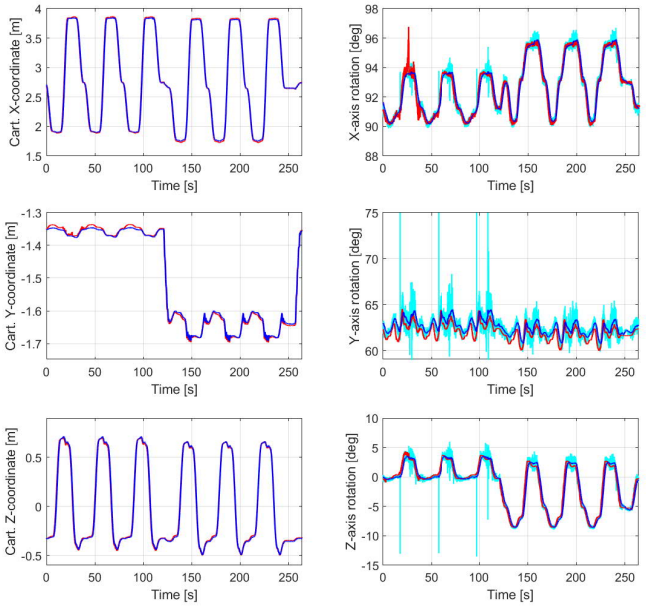

Fig. 9. Comparison of the optically measured 6 DOF TCP poses (with ICP calibration). The red lines are associated with SLAM, the blue lines represent the marker-based tracking, and the light blue orientations denote the unconditioned signals.

combining marker-based tracking with SLAM pose estimation. The test case using a 6 DOF hydraulic manipulator, with a reach of approximately 5 m, assumed equal availability (same frequency) of the SLAM and marker poses. However, it was shown that the SLAM camera, located near the TCP, provided higher quality orientation measurements in relation to the marker-based orientation measurements. In reality, it is foreseen that the SLAM module is required to do the majority of the work in the TCP pose estimation, as the distances between the TCP and the base of the manipulator are quite large in actual mining manipulators, which will degrade the accuracy of any marker-based tracking system. In addition, occlusions will be a challenge in mining manipulators that can rotate approximately 360°. Thus, marker-based tracking is likely to be more useful as a secondary sensor module, which can be realized, for example, with the CWA data fusion method by tuning the weights appropriately. The utilization of marker-based tracking for calibration, and for example, condition monitoring, should be explored in the

TABLE IV

MEAN ABSOLUTE ERRORS IN EACH CASE.

	Fig. 6	Fig. 7	Fig. 8	Fig. 9
x [m]	0.0415	0.0262	0.0292	0.0216
y [m]	0.0052	0.0046	0.0044	0.0046
z [m]	0.0210	0.0081	0.0095	0.0142
γ_x [deg]	1.1221	1.0961	1.0924	0.2775
γ_y [deg]	0.6462	0.6429	0.6079	0.6423
γ_z [deg]	1.8541	1.8277	1.8171	0.6104

TABLE V

MAXIMUM ABSOLUTE ERRORS IN EACH CASE.

	Fig. 6	Fig. 7	Fig. 8	Fig. 9
x [m]	0.1195	0.0615	0.0619	0.1457
y [m]	0.0288	0.0163	0.0214	0.0174
z [m]	0.0894	0.0318	0.0347	0.0815
γ_x [deg]	3.4733	3.5547	3.4241	3.1852
γ_y [deg]	1.8698	2.1864	1.3959	1.9344
γ_z [deg]	5.3777	7.1867	6.0092	2.7358

future.

Although the two visual sensor modules produced seemingly high performance, the challenge lies in the numerous uncertainties present in the system. These follow especially from the calibration that, even in the case of offline data analysis, resulted in relatively considerable errors. Furthermore, a new calibration method is required for future online experiments.It is also perceived that the encoder setup used for reference measurements may provide the least accurate TCP pose measurement. Being able to match the visual pose measurements to the analytical TCP pose is desirable in the sense that to control the manipulator, knowledge of the joint states is usually required, which could be achieved by using an inverse kinematic model. In addition, although the applied kinematic model is based on the rigidity assumption, these types of long-reach manipulators are very flexible due to their length and high payload-to-own-weight ratio. This flexibility and the following non-rigid kinematics are also a central research problem related to the TCP pose estimation in long-reach manipulators, and solving this with external sensors is a long-term goal of this research. Thus, alternative control methods that are not directly based on the joint states should be pursued.

ACKNOWLEDGMENT

This work was supported by the Doctoral School of Industry Innovations (DSII) of Tampere University. This work was also carried out with the support of Centre for Immersive Visual Technologies (CIVIT) research infrastructure, Tampere University, Finland.

REFERENCES

[1] P. Corke, J. Roberts, and G. Winstanley, "Vision-based control for mining automation," *IEEE Robot. Automat. Mag.*, vol. 5, no. 4, pp. 44–49, 1998.

[2] J. Vihonen, J. Mattila, and A. Visa, "Joint-space kinematic model for gravity-referenced joint angle estimation of heavy-duty manipulators," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3280–3288, 2017.

[3] P. Mäkinen, M. M. Aref, J. Mattila, and S. Launis, "Application of simultaneous localization and mapping for large-scale manipulators in unknown environments," in *Cybernetics and Intelligent Systems (CIS) and IEEE Conf. Robotics, Automation and Mechatronics (RAM), 2019 IEEE 9th Inter. Conf.* IEEE, 2019.

[4] G. Flandin, F. Chaumette, and E. Marchand, "Eye-in-hand/eye-to-hand cooperation for visual servoing," in *Proc. 2000 ICRA. Millennium Conf. IEEE Int. Conf. Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, vol. 3. IEEE, 2000, pp. 2741–2746.

[5] V. Lippiello, B. Siciliano, and L. Villani, "Eye-in-hand/eye-to-hand multi-camera visual servoing," in *Proc. 44th IEEE Conf. Decision and Control*. IEEE, 2005, pp. 5354–5359.

[6] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, 2017.

[7] P. Cheng, B. Oelmann, and F. Linnarsson, "A local positioning system for loader cranes based on wireless sensors—a feasibility study," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 8, pp. 2881–2893, 2011.

[8] 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," *Inter. J. Robot. Res.*, vol. 29, no. 10, pp. 1338–1352, 2010.

[9] 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," *Autom. Construction*, vol. 65, pp. 51–64, 2016.

[10] C.-J. Liang, K. M. Lundeen, W. McGee, C. C. Menassa, S. Lee, and V. R. Kamat, "A vision-based marker-less pose estimation system for articulated construction robots," *Autom. Construction*, vol. 104, pp. 80–94, 2019.

[11] T. Kivelä, J. Mattila, J. Puura, and S. Launis, "Redundant robotic manipulator path planning for real-time obstacle and self-collision avoidance," in *Int. Conf. Robotics in Alpe-Adria Danube Region*. Springer, 2017, pp. 208–216.

[12] W. Elmenreich, "Fusion of continuous-valued sensor measurements using confidence-weighted averaging," *J. Vib. Control*, vol. 13, no. 9-10, pp. 1303–1312, 2007.

[13] 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.

[14] S. W. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 42, no. 1, pp. 97–101, 2000.

[15] R. N. Jazar, *Theory of Applied Robotics - Kinematics, Dynamics, and Control*. Dordrecht, the Netherlands: Springer, 2010.