2022 The 22st International Conference on Control, Automation and Systems (ICCAS 2022)
BEXCO, Busan, Korea, Nov. 27∼Dec. 01, 2022

Model-Free Sensor Fusion for Redundant Measurements Using Sliding Window Variance

P. Mäkinen[1*], P. Mustalahti[1], S. Launis[2] and J. Mattila[1]

[1]Automation Technology and Mechanical Engineering, Tampere University, Tampere, Finland,
(petri.makinen@tuni.fi, pauli.mustalahti@tuni.fi, jouni.mattila@tuni.fi) * Corresponding author
[2]Rock Technologies and Drilling, Sandvik Mining and Construction Oy,
Tampere, Finland (sirpa.launis@sandvik.com)

Abstract: In this paper, a model-free data fusion method for combining redundant sensor data is presented. The objective is to maintain a reliable tool center point pose measurement of a long-reach robotic manipulator using a visual sensor system with multiple cameras. The fusion method is based on weighted averaging. The weight parameter for each variable is computed using the sliding window variance with N latest observations. After each sliding window, the window length N is updated, and simple transition smoothing is included. For experimental validation, two sets of pose trajectory data from redundant visual sensors were obtained: 1) using a camera located near the tip of a long-reach manipulator running a simultaneous localization and mapping (SLAM) algorithm and 2) marker-based tracking with cameras located near the base of the manipulator. For pose tracking, a fiducial marker was attached near the SLAM camera. The proposed methodology was examined using a real-time measurement setup and offline data analysis using the recorded data. The results demonstrate that the proposed system can increase the overall robustness and fault tolerance of the system, which are desired features for future autonomous field robotic machines.

Keywords: Sensor Fusion, Machine Vision, Sensor Systems and Applications

1. INTRODUCTION

The heavy machinery industry is taking major leaps toward electrification and autonomous systems. These heavy-duty mobile machines require new intelligent algorithms and sophisticated sensors [1] in order to work independently in harsh environments, such as mines. A variety of long-reach robotic manipulators are found in such machines to perform various tasks related to mining and heavy lifting, for example [2]. One of the key challenges is replacing human vision and decision making with sensors and computerized algorithms in order to perform work tasks autonomously. For this purpose, measurement information about the manipulator's tool center point (TCP) is typically required. The TCP pose (3 degrees-of-freedom (DOF) position and 3 DOF orientation) can be obtained using forward kinematics with joint encoders. For a small-scale, ideal industrial robot, an accurate forward kinematic model can be obtained. However, this is not the case for long-reach manipulators, as they have significant structural flexibilities that are not considered by traditional rigid-body kinematics. Consequently, visual servoing methods have been well established for small-scale industrial robots. In contrast, for long-reach manipulators working in unstructured environments, there are challenges with visual sensing related to camera calibration, view distance, field of view, and occlusions, for example [3].

In an attempt to replace human vision in applications striving toward autonomy, a wide variety of visual sensors have been investigated, including radar technology and optical methods, such as laser scanners and camera systems [4]. Data provided by different proprioceptive and exteroceptive sensors can be combined to obtain a more accurate or robust picture of an observed system. This process of combining sensor information is called multi-sensor fusion, or simply sensor fusion. Based on how sensor information is utilized, fusion methods are usually classified as competitive, complementary, or cooperative systems [5]. For sensor fusion, the Kalman filter and its nonlinear variants are popular methods [6-7]. Neural networks and fuzzy set theory have also been investigated [8-9]. The most mature branch of sensor fusion is perhaps related to self-driving vehicles, which have been avidly examined [10-11] and with, for example, Tesla Autopilot available for consumer vehicles. These systems are built on deep learning algorithms, requiring massive amounts of training data, which, scale-wise, are not feasible for the low-volume heavy machinery industry.

One of the previous studies on multi-sensor integration [5] argued that the key to intelligent fusion of disparate sensory information is to provide an effective model of sensor capabilities. However, in some cases, finding a sufficient sensor model may not be possible. Research on such model-free sensor data fusion methods is very limited and restricted to simple, albeit potentially effective, methods. For example, in [12], fusion was carried out using confidence weighted averaging. However, determining the confidence functions for weight computations in dynamic, online scenarios has not been well established.

In this paper, we focus on combining continuous measurements of the same variables from different sensors in an attempt to increase the system's robustness and reliability. The fusion method is a statistical approach based on confidence weighted averaging [12], with our contributions including determining the weight parameters in a dynamic manner using sliding window variance (sample variance) instead of a specified confidence function. The sliding window length is updated after each individ-

ual window, and a simple approach for transition smoothing is also presented. The proposed methodology was investigated using a real-time setup comprising a long-reach hydraulic manipulator. The objective was to estimate the end effector's pose with visual sensors [13]. The first pose estimate was obtained using a camera running a simultaneous localization and mapping (SLAM) algorithm, with the camera attached near the tip of the manipulator. The second pose estimate was obtained using marker-based tracking, with a fiducial marker attached near the SLAM camera. Before sensor fusion was performed, the pose variables were extrinsically calibrated to a concurrent coordinate system according to [14].

The underlying motivation with this configuration is that the SLAM algorithm has a narrow but accurate view, whereas the marker-tracking cameras are placed on top of a machine and provide a wider view but a less accurate pose measurement. A conceptual example is shown in Fig. 1. For this application, a marker provides a fixed and repeatable target. Consequently, the visual sensors are able to complement each other, but also provide the necessary redundancy, as both sensing methods are susceptible to faulty situations. These include, for example, marker occlusions and insufficient feature extraction for SLAM.

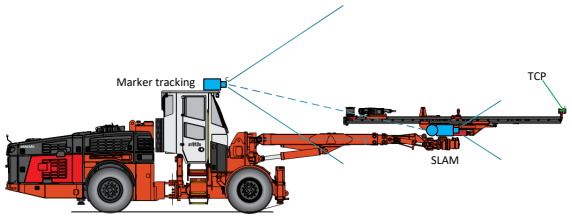

Fig. 1. The overall conceptual design: The TCP is observed using a visual sensor system comprising marker tracking and SLAM modules. The Sandvik DT912D single boom tunneling jumbo is shown as an example.

The remainder of the paper is organized as follows: The data fusion methodology is described in Section 2, the experimental setup is detailed in Section 3, the results are discussed in Section 4, and finally Section 5 concludes the paper.

2. METHODOLOGY

2.1 Data fusion using sliding window variance

A fused sensor signal can be formulated by taking the weighted average of all the sensor signals that estimate the same variable:

$$x_F = \sum_{i=1}^{n} w_i x_i, \tag{1}$$

where x_F is the fused signal, w_i denotes the weight parameters, x_i denotes the redundant sensor data, and n is the number of sensors. The fused variance can be written as

$$\sigma_F^2 = \sum_{i=1}^{n} w_i \sigma_i^2, \tag{2}$$

where σ_F^2 is the fused variance, and σ_i^2 denotes the input signal variance.

To obtain the optimal fused measurement, the weight parameters should be chosen so that the fused variance is minimized, which can be achieved by solving the following minimization problem:

$$\arg\min_{w_i} \sum_{i=1}^{n} w_i^2 \sigma_i^2, \tag{3}$$

with the sum of all weights w_i equal to 1. Solving the minimization problem results in the following equation to compute the weights:

$$w_i = \frac{1}{\sigma_i^2} \sum_{j=1}^{n} \frac{1}{\sigma_j^2}. \tag{4}$$

Substituting Eq. (4) into Eq. (2) shows that for $n \geq 2$, the fused variance is always smaller than the input variances.

The variance for a given measurement signal is computed over a sliding window of length N data points:

$$\sigma_i^2 = \frac{1}{N-1} \sum_{j=1}^{N} |x_j - \mu_i|^2, \tag{5}$$

where μ_i denotes the mean of x_j over the sliding window of N observations and is computed as

$$\mu_i = \frac{1}{N} \sum_{i=j}^{N} x_j. \tag{6}$$

Then, the sliding window variances are used to compute the weight parameters using Eq. (4) for data fusion.

The rationale is that computing the weights based on sliding window variances of redundant measurements will emphasize better-quality signals, as it is expected that a signal with less variance is more accurate. This derives from the assumption that the measurements are reliable, and grossly faulty measurements are detected and discarded before the data fusion procedure.

2.2 Updating the sliding window length

The length of the sliding window, denoted by N, is updated at the end of each window, which is conducted as follows:

$$N = k_N \max|\boldsymbol{\mu}_{x_1} - \boldsymbol{\mu}_{x_j}|, \tag{7}$$

where $\boldsymbol{\mu}_{x_1}$ and $\boldsymbol{\mu}_{x_j}$, $j \in \{2, ..., n\}$ denote vectors of the mean values of each sensor measurement, computed over the entire current sliding window. The largest absolute mean difference is used to compute the next window length, and the coefficient k_N is used to tune the window length to a desired scale. Note that N is rounded to an integer value. The sliding window length should be constrained between the minimum N_{min} and maximum N_{max} values to avoid computational issues.

2.3 Transition smoothing

Raw sensor data can contain occasional outliers, or some sensors may cease to operate, which requires smoothly and safely transitioning the fused sensor signal to exclude the unavailable sensor measurement. For this purpose, we use a simple transition smoothing method. The fused sensor signal $\hat{x}_F$ is computed using the following condition:

$$\hat{x}_{F,new} = \begin{cases} x_F & \text{if } |\hat{x}_{F,previous} - \hat{x}_{F,current}| < \epsilon \\ x_{F_{corr}} & \text{otherwise} \end{cases}. \tag{8}$$

The error coefficient ϵ determines the limit after which the transition smoothing is applied. If the absolute difference between the previous and current fused values is less than the designated error coefficient, the next fused value is computed normally using Eq. (1). If the condition is not met, the next fused value is predicted in a naïve manner by using linear interpolation. A polynomial function $p(x)$ of k degree is written as:

$$p(x) = p_1 x^k + p_2 x^{k-1} + ... + p_k x + p_{k+1}. \tag{9}$$

Considering a first-order polynomial, the linear system can be presented as follows:

$$\begin{bmatrix} t_1 & 1 \\ t_2 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \hat{x}_{F,previous} \\ \hat{x}_{F,current} \end{bmatrix}, \tag{10}$$

where $\{t_1, t_2\}$ are time stamps dictating the rate of the desired transition smoothing, and $\{p_1, p_2\}$ are polynomial coefficients to be solved. Then, the new corrected fused value for the next time step is obtained using Eq. (9):

$$x_{F_{corr}} = p_1 * (t_1 + Ts) + p_2, \tag{11}$$

where Ts is the sampling period.

3. EXPERIMENTAL SETUP

For validating the proposed model-free sensor fusion pipeline, two sets of 6-DOF pose trajectories were obtained from two redundant visual sensing methods. The experimental system is illustrated in Fig. 2, and it comprised a hydraulic manipulator, a stereo camera for SLAM, and a motion capture system for marker tracking.

The real-time system controlling the hydraulic manipulator was a Beckhoff CX2030 industrial PC. During the experiments, the manipulator was moved arbitrarily around its workspace. All the measurement data were collected by the Beckhoff PC to ensure time synchronization.

3.1 SLAM module

A ZED2 stereo camera was attached near the tip of the hydraulic manipulator. The camera was connected to a dedicated Linux PC running ROS (the robot operating system), and 720p images were published using the manufacturer-provided ROS node.

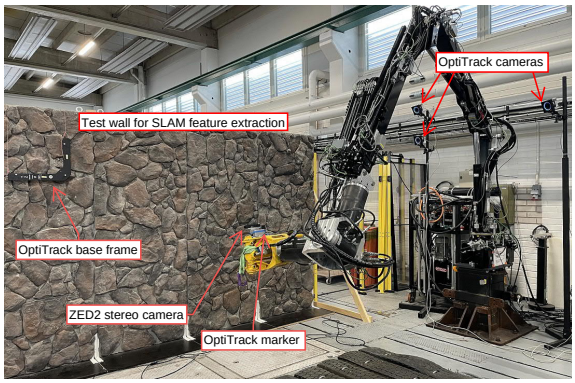

Fig. 2. The experimental setup: The hydraulic manipulator was moved arbitrarily around its workspace, and two sets of 6-DOF pose trajectory data were obtained using 1) the ZED2 stereo camera for SLAM and 2) the OptiTrack motion capture system for tracking the marker pose with respect to the OptiTrack L-frame.

For SLAM, we used the open-source ORB-SLAM2 Stereo[1] algorithm [15]. The algorithm ran on the dedicated Linux PC in real time using the images published by the ZED2 ROS node, and the 6-DOF pose trajectory data were transmitted to the Beckhoff industrial PC via UDP (user datagram protocol).

3.2 Marker tracking module

The marker-tracking module comprised three OptiTrack Prime 17W wide angle coverage cameras, a passive marker, and a base frame. The cameras were placed on high pillars around the base of the manipulator. The base frame (or OptiTrack L-frame) was placed in view of the cameras, and the marker was attached near the tip of the manipulator. The system then tracked the 6-DOF marker pose with reference to the L-frame.

A dedicated laptop with OptiTrack's Motive software was used to set up the marker-tracking module. A MATLAB plugin was configured to transmit the 6-DOF pose trajectory data to the Beckhoff industrial PC.

3.3 Signal calibration

Sensor fusion requires variables that represent the same information. A requirement before fusion is that the measured variables are transformed from each sensor's local coordinate system to a common one [16]. For pose estimates, this implies that the poses must be expressed with respect to a concurrent coordinate system. This extrinsic calibration is defined as a rigid transformation, comprising a rotation matrix and a translation vector, from one coordinate system to another.

Obtaining this rigid relationship can be a challenging task especially with field robotic systems due to the unstructured and unknown environments. In this work, we used a probabilistic point set matching-based methodology [14] to find the transformation between the two visual sensor coordinate systems. Specifically, the SLAM poses

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

were calibrated to the OptiTrack's base frame.

3.4 Real-time implementation

After the extrinsic calibration, the two pose measurements were expressed in the same coordinate system and fused according to the methodology detailed in Section 2. The error limit for transition smoothing in Eq. (8) was computed using the maximum absolute errors resulting from the extrinsic calibration as follows: $\epsilon = k_\epsilon \epsilon_{calib}$, where k_ϵ is a multiplication factor used to tune the transition smoothing. The relevant parameters applied in the experiments are shown in Table 1 and they were manually optimized for the investigated application.

Table 1. Applied parameters.

N_{min}	N_{max}	k_N	k_ϵ	t_1	t_2	Ts
200	4000	$30 * 10^4$	1.0	0 s	0.5 s	1 ms

The data fusion algorithm was initialized using the set maximum window size, which took 4 s with the applied parameters. Normal operation was commenced only after the initialization. The sliding window was implemented so that the overall window size was constant (the set maximum size), with the unused elements set to zero. The variance and mean value Eqs. (5)–(6) related to each measured variable were computed over the nonzero elements, with the length of the nonzero variables depending on the current window size N.

In the case of noisy data, we suggest using a geometric moving average filter [17] before data fusion for improved signal quality. The filter is formulated as follows:

$$X_i = (1 - \alpha) X_{i-1} + \alpha x_i, \quad i > 0, \tag{12}$$

where X_i is the conditioned output signal at time i, x_i is the unconditioned input signal at time i, and $0 < \alpha \leq 1$ is the filter gain, for which a low value is advised. The results presented in this paper, however, were unfiltered.

4. RESULTS AND DISCUSSION

The data fusion algorithm was tested online on the real-time system, and data were recorded for further offline data analysis. The results presented here were obtained in MATLAB's Simulink environment using the recorded data. Three cases were studied: Case 1: normal operation, Case 2: updating the sliding window length, and Case 3: transition smoothing.

For Case 1, Figs. 3–6 illustrate the poses, weight parameters, variances, and sliding window lengths obtained from the same measurement. In general, the red lines denote SLAM signals, the green lines denote marker-tracking signals, and the black lines denote fused signals. Fig. 3 shows the poses for which the positional components of both measurements perform similarly. Regarding the orientation measurements, the marker-tracking signals were noisy, whereas SLAM provided better-quality signals. Thus, the fused signals emphasized the orientations provided by the SLAM module. The difference between the position and orientation measurements is also demonstrated in the weight parameters in Fig. 4: The position signals had similar qualities, resulting in uniform weight parameter distributions. The SLAM orientation measurements had better qualities in the sense that the sliding window variances were smaller, resulting in larger weight parameters for the SLAM orientations. Note that the weight parameters were set to zeros during the initialization, but during normal operation the total sum of the weights is equal to 1. The respective signal variances are shown in Fig. 5. As discussed, the fused variances are always smaller than the input variances as a result from Eq. (3). Finally, Fig. 6 illustrates the sliding window lengths for each variable. As shown, the fusion algorithm was initialized with the set maximum window length $N_{max} = 4000$, after which the lengths were updated after each sliding window according to Eq. (7).

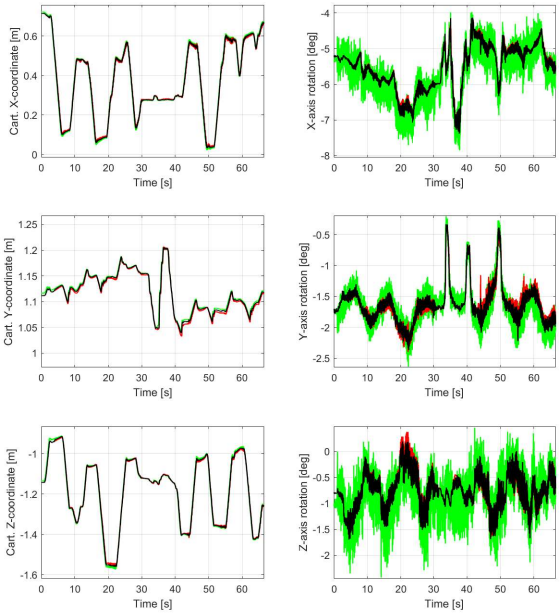

Fig. 3. Case 1: All six pose variables are shown. The red lines are the SLAM signals, the green lines are the marker-tracking signals, and the black lines are the fused signals.

For Case 2, the goal was to demonstrate the impact of the sliding window length on the fusion output. Figs. 7–9 show the fused poses for three instances: constant (minimum) sliding window length, constant (maximum) sliding window length, and variable sliding window length (same as in Case 1). For clearer visualization, at 30 s, the SLAM position signals were artificially increased by 0.15 m, and the SLAM orientation signals were increased by 5°. The left figures show the pose signals, and the right figures show the respective sliding window lengths. On the left, the red lines are the SLAM signals, the green lines are the marker-tracking signals, and the black lines are the fused signals. Fig. 7 shows the results when the set minimum sliding window length of $N_{min} = 200$ is

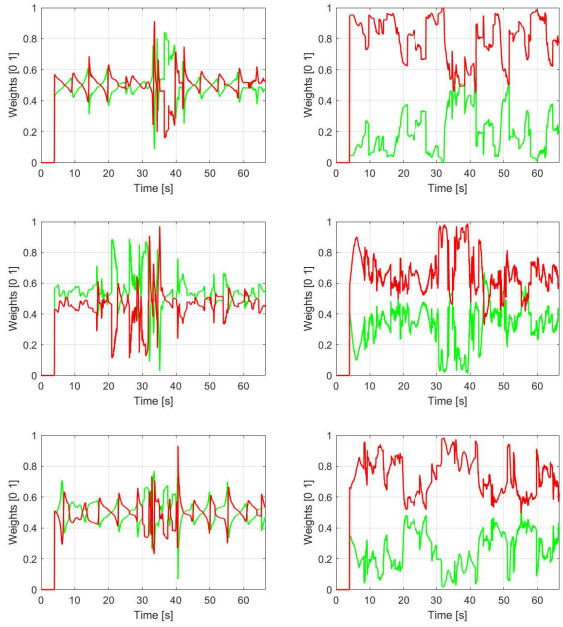

Fig. 4. Case 1: The weight parameters for each fused signal at the given time stamps. The red lines represent the SLAM weight parameters, whereas the green lines represent the marker-tracking weight parameters.

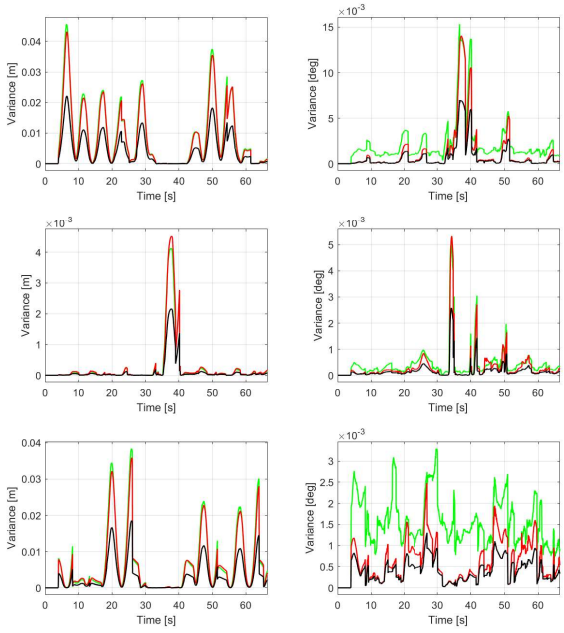

Fig. 5. Case 1: Computed variances over the sliding window for each signal. The red lines are the SLAM signal variances, the green lines are the marker-tracking signal variances, and the black lines are the fused variances.

used. Applying a small window length for signals with approximately equal variances results in poor fused signal quality, when the difference between the input signals increases. This is shown in the position signals. For orientation signals with clearly different variances (due to the noise level), the fused output strictly emphasizes

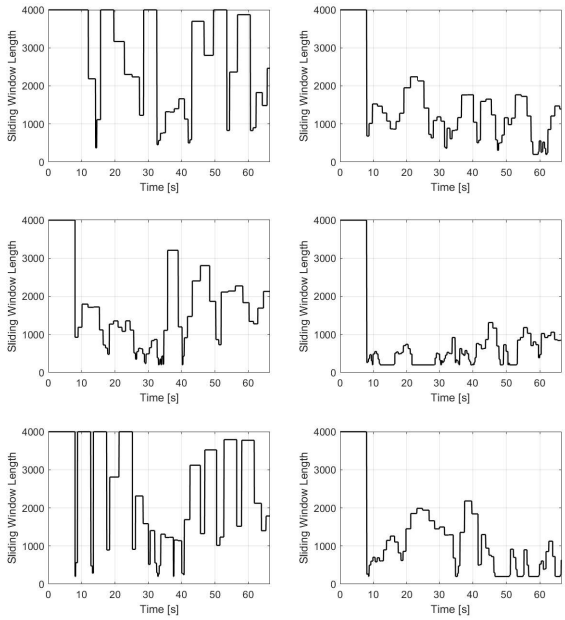

Fig. 6. Case 1: The sliding window lengths for each signal at the given time stamps.

the better-quality SLAM orientation signal. However, the small window length still induces some of the noise in the fused output. Fig. 8 shows the same results, while the set maximum window length of $N_{max} = 4000$ is used. In this case, the resulting fused signals are more restrained. Finally, Fig. 9 shows the results for the variable sliding window lengths. As illustrated, the window lengths first float between the set minimum and maximum values. However, after the SLAM signals are artificially increased, the window lengths jump to the set maximum values. The difference compared with the previous case of using the set maximum window length is that with variable lengths the computations can be performed over a smaller number of elements. Moreover, as demonstrated in Figs. 7–9 before the 30 s marks, the fused signals are slightly better with smaller window lengths due to the small difference between the input measurements. Thus, the absolute mean difference between the signals, computed over the current sliding window, was chosen as the basis for updating the window length in Eq. (7).

For Case 3, the aim was to assess the fusion algorithm's performance when a measured signal is lost, and transition smoothing is required. Smoothing is applied to avoid sudden, undesired changes with large amplitudes in the fused output signal. The results are illustrated in Fig. 10, in which, at 30 s, the SLAM position signals were again artificially increased by 0.15 m, and the SLAM orientation signals were increased by 5°, respectively. Then, the SLAM measurement was switched off, after which the fusion algorithm switched to utilize only the marker tracking-based measurement. Then, the SLAM measurement was switched back on, and the fusion algorithm resumed to utilize both pose measurements. For comparison, the same procedure was repeated by switching the marker tracking off and on. The red lines are the

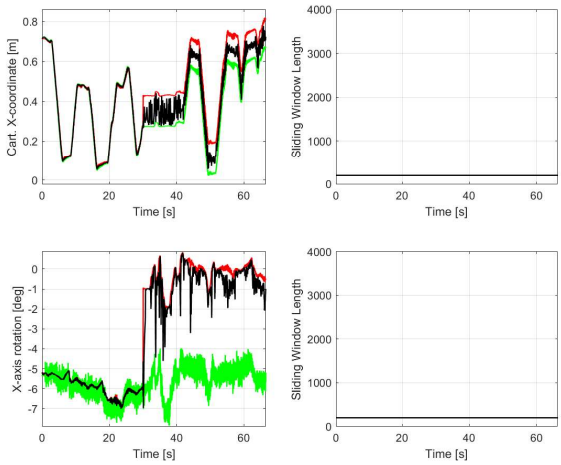

Fig. 7. Case 2: A constant sliding window length of 200 was used: At 30 s, the SLAM position signals were artificially increased by 0.15 m, and the SLAM orientation signals were increased by 5°. The left figures show the signals, and the right figures show the respective sliding window lengths. On the left, the red lines are the SLAM signals, the green lines are the marker-tracking signals, and the black lines are the fused signals.

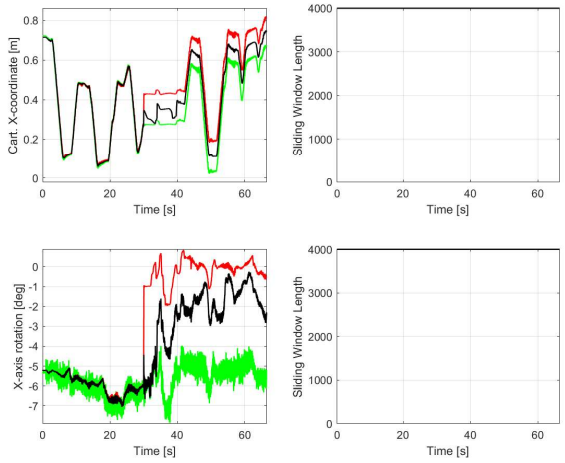

Fig. 8. Case 2: A constant sliding window length of 4000 was used: At 30 s, the SLAM position signals were artificially increased by 0.15 m, and the SLAM orientation signals were increased by 5°. The left figures show the signals, and the right figures show the respective sliding window lengths. On the left, the red lines are the SLAM signals, the green lines are the marker-tracking signals, and the black lines are the fused signals.

SLAM signals, the green lines are the marker-tracking signals, the black lines are the fused signals with transition smoothing, and the magenta lines are the fused signals without transition smoothing. The switching time $t_2 - t_1$ in Eq. (10) was 0.5 s, which dictated the desired convergence time toward the available pose measurement, and the "jump" occurs when the error coefficient ϵ in Eq. (8) is reached. As shown, the transitions

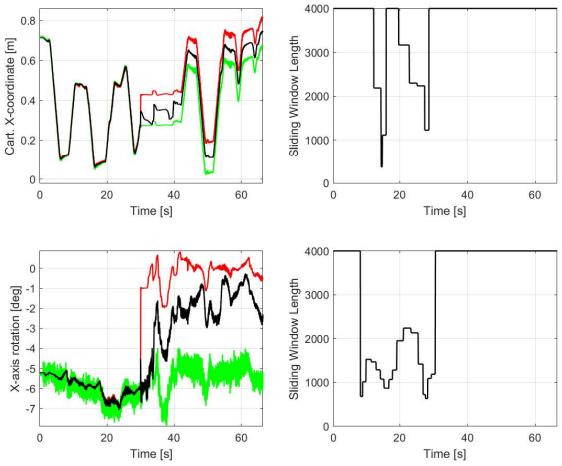

Fig. 9. Case 2: Variable sliding window lengths were used: At 30 s, the SLAM position signals were artificially increased by 0.15 m, and the SLAM orientation signals were increased by 5°. The left figures show the signals, and the right figures show the respective sliding window lengths. On the left, the red lines are the SLAM signals, the green lines are the marker-tracking signals, and the black lines are the fused signals.

are appropriately smoothed, and the effectiveness can be tuned by adjusting the parameters.

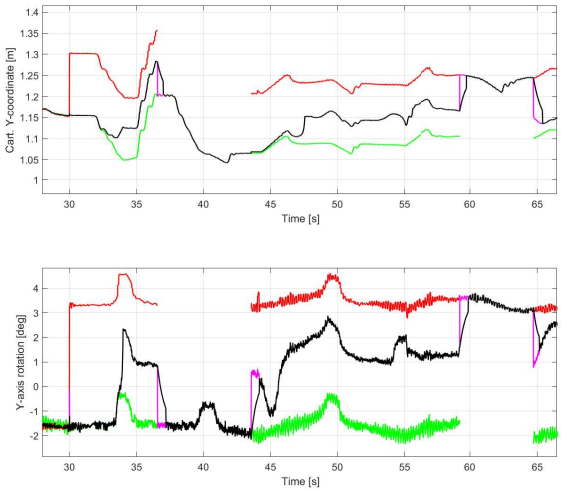

Fig. 10. Case 3: At 30 s, the SLAM position signals were artificially increased by 0.15 m, and the SLAM orientation signals were increased by 5°. The transitioning of the fused signal is demonstrated when the other measurement signal is lost. The red lines are the SLAM signals, the green lines are the marker-tracking signals, the black lines are the fused signals with transition smoothing, and the magenta lines show the fused signals without transition smoothing.

The error coefficients have to be carefully set, as values that are too low will have a deteriorating effect on the fusion output; the transition smoothing should enable only when sudden, undesired changes occur in the input measurements. However, values that are too large will

render the smoothing ineffective.

5. CONCLUSION

In this paper, we examined the problem of directly fusing continuous sensor data in a real-time setting. The presented model-free pipeline is a statistical approach based on weighted averaging, in which the weight parameters are constantly updated using the sliding window variances of the respective signals to be fused. A method for updating the window length was shown, along with a simple transition smoothing design.

Results based on real-time experiments were presented: 6-DOF pose trajectory data from two independent, redundant visual sensors were fused in an optimal manner in the sense that the variances of the fused signals were minimized with respect to the input variances, computed over the current sliding windows. The experimental results demonstrated that the proposed methodology can increase the system's robustness and fault tolerance, which are the desired features for future autonomous field robotic machines.

Some challenges of this methodology include the lack of a model, which makes the system rely more on sophisticated sensor self-diagnostics before fusion occurs, so that faulty measurements are detected and discarded before fusion is executed.

ACKNOWLEDGEMENT

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 the Centre for Immersive Visual Technologies (CIVIT) research infrastructure, Tampere University, Finland.

REFERENCES

[1] T. Machado, D. Fassbender, A. Taheri, D. Eriksson, H. Gupta, A. Molaei, P. Forte, P. K. Rai, R. Ghabcheloo, S. Mäkinen, A. J. Lilienthal, H. Andreasson, and M. Geimer, "Autonomous heavy-duty mobile machinery: A multidisciplinary collaborative challenge," in *2021 IEEE International Conference on Technology and Entrepreneurship (ICTE)*, pp. 1–8, 2021.

[2] L. Lopes, T. Miklovicz, E. Bakker, and Z. Milosevic, "The benefits and challenges of robotics in the mineral raw materials sector-an overview," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1507–1512, 2018.

[3] J. A. Marshall, A. Bonchis, E. Nebot, and S. Scheding, "Robotics in mining," in *Springer handbook of robotics*, Springer, pp. 1549–1576, 2016.

[4] J. Kocić, N. Jovičić, and V. Drndarević, "Sensors and sensor fusion in autonomous vehicles," in *2018 26th Telecommunications Forum (TELFOR)*, IEEE, pp. 420–425, 2018.

[5] H. F. Durrant-Whyte, "Sensor models and multisensor integration," in *Autonomous Robot Vehicles*, Springer, New York, pp. 73–89, 1990.

[6] S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion Kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, 2004.

[7] J. Gross, Y. Gu, S. Gururajan, B. Seanor, and M. Napolitano, "A comparison of extended Kalman filter, sigma-point Kalman filter, and particle filter in GPS/INS sensor fusion," in *AIAA Guidance, Navigation, and Control Conference*, p. 8332, 2010.

[8] E. Bostanci, B. Bostanci, N. Kanwal, and A. F. Clark, "Sensor fusion of camera, GPS and IMU using fuzzy adaptive multiple motion models," *Soft Computing*, vol. 22, no. 8, pp. 2619–2632, 2018.

[9] S. Yazdkhasti and J. Z. Sasiadek, "Multi sensor fusion based on adaptive Kalman filtering," in *Advances in Aerospace Guidance, Navigation and Control*, Springer, Cham, pp. 317–333, 2018.

[10] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, p. 2140, 2021.

[11] J. Fayyad, M. A. Jaradat, D. Gruyer, and H. Najjaran, "Deep learning sensor fusion for autonomous vehicle perception and localization: A review," *Sensors*, vol. 20, no. 15, p. 4220, 2020.

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

[13] P. Mäkinen, P. Mustalahti, S. Launis, and J. Mattila, "Redundancy-based visual tool center point pose estimation for long-reach manipulators," in *2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 1387–1393, 2020.

[14] P. Mäkinen, P. Mustalahti, S. Launis, and J. Mattila, "Probabilistic camera-to-kinematic model calibration for long-reach robotic manipulators in unknown environments," in *2022 IEEE 17th International Conference on Advanced Motion Control (AMC)*, pp. 48–55, 2022.

[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] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

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