Novel Adaptive Sensor Fusion Methodology for Hand Pose Estimation with Multi-Leap Motion

Salih Ertug Ovur, *Student Member, IEEE,* Hang Su, *Member, IEEE,* Wen Qi, *Student Member, IEEE,* Elena De Momi, *Senior Member, IEEE,* and Giancarlo Ferrigno, *Senior Member, IEEE*

***Abstract*—In the past decade, touchless interaction with objects drawn increasing attention in the wide range of applications from entertainment to the real-time control of robots. In this aim, many devices such as Leap Motion and Microsoft Kinect developed for tracking hand posture. However, successful realization of these sensors for real-time touchless interaction applications still needs to be improved. Subsequently, in this paper, adaptive sensor fusion methodology is proposed for hand pose estimation with two Leap Motions. Developed adaptive methodology is capable of performing stable and continuous hand position estimation in real-time even when a single sensor is unable to detect a hand. Self-calibration algorithm is implemented to tolerate the incompatibility in sensor reference frames. Two separate Kalman filters are adopted for adaptive sensor fusion of palm position and orientation. Proposed adaptive sensor fusion method is verified with various experiments in six degrees of freedom in space.**

***Index Terms*—Adaptive Sensor Fusion, Kalman Filter, Leap Motion.**

I. Introduction

IN recent decades, technological innovations have removed the perception that human-computer interaction requires physical contact and have introduced the methodology of touchless interaction [1]. For touchless interaction, the key-point is the estimation of the hand pose that includes both position and orientation of the hand frame [2]. Nowadays, there are various instruments available in the commercial market such as Leap Motion Controller (LMC[1]) and Microsoft Kinect[2] that allows contactless interaction in the human-computer interaction (HCI) [3], [4], [5].

Applications already spreaded to various fields such as virtual reality (VR) surgeries [6], robot-manipulation [7] and sign language recognition [8]. Touchless teleoperation of the RAVEN-II surgical robot is performed by using LMC in [9]. In the literature, not only the master-slave based interaction is considered, but also haptic feedback is introduced in [10], by using LMC and wearable tactile devices which is realized in VR application. Thanks to the non-physical sterile interaction, applications have also extended to the clinical field [11]. Another HCI application held for medical image data manipulation during the surgeries by using LMC realized in [12]. Moreover, LMC is frequently used for hand gesture recognition as we did in our previous works, a novel method proposed by using LMC sensor to classify electromyography signals for hand gesture recognition [13].

Almost all these papers in the literature agreed on that, due to the occlusion problem, direct use of LMC is not applicable and reliable in most cases. Therefore, multiple sensors become a solution for continuous and smooth data flow with sensor fusion [14]. Sensor fusion is possible with the variety of sensors such as LMC-Microsoft Kinect [15], LMC-Myo Armband [16], LMC-LMC [17]. In order to make sensor fusion, Bayesian state estimators such as particle filter or kalman filter used frequently in the literature [18].

Literature research showed that there is still a room for improvements due to the lackness of stable and continous method for touchless interaction. Therefore, in order to overcome these problems, a more stable and accurate touchless interface is proposed in this paper. This paper puts forth a novel method for estimating hand pose with two LMCs. Novel adaptive sensor fusion with two kalman filters proposed in order to estimate position and orientation of the palm center. The greatest strength of our method is that even if occlusion occurs in one sensor, it can maintain smooth estimation. Moreover, measurement covariance is adaptively changed in order to endeavour with the changes in the environment. The contribution of this paper includes:

- Thanks to the proposed adaptive sensor fusion methodology, smooth and continuous palm position and orientation estimation, even one LMC does not detect the hand.
- Real-time execution is performed without any preprocessing. Synchronization of sensors is performed via ROS.
- Self-calibration algorithm automatically registers the sensors to the same reference frame. Arbitrary positioning of sensors is compensated.

In the following sections, the design, working and experimental results of the proposed Sensor fusion method are discussed. In Section II, various types of sensor fusion methods are briefly described. In Section III, a brief description of the two sensors under study is given. In Section IV, basic Kalman filter is discussed from sensor fusion point of view. The model equations and parameters of the Kalman filter for the problem of finger tip position tracking are derived. In Section V, the experimental setup is explained and the results are discussed which is followed by remarks and conclusions

*This work was supported by the European Commission Horizon 2020 research and innovation program, under the project SMARTsurg, grant agreement No. 732515.

Salih Ertug Ovur, Hang Su, Wen Qi, Elena De Momi and Giancarlo Ferrigno are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133, Milano, Italy. {salihertug.ovur@mail.polimi.it} {hang.su, wen.qi, elena.demomi,giancarlo.ferrigno }@polimi.it

[1]Leap Motion, https://www.leapmotion.com/

[2]Microsoft Kinect, https://developer.microsoft.com/en-us/windows/kinect

in the subsequent section.

II. Related Works

Sensor fusion with LMC attracted variety of researches. An adaptive human-robot interface adopted in [19] by using kalman filter to estimate orientation and particle filter to estimate palm position of human hands. Proposed method used to operate dual robots by using double hands. In the following research [20] deployed from the same authors, they introduced markerless human-robot interface by using interval kalman filter to estimate palm position and improved particle filter to estimate the orientation of the human hands. Thanks to this new algorithm they improved accuracy of palm position and stability of the orientation estimation with respect to their previous work.

Apart from use of single LMC, due to the occlusion or limited workspace problem of LMC, multiple sensors are also considered in the literature. [16] is introduced sensor fusion of Myo Armband - LMC sensors in order to improve the performance and also to have full model of arm motion including forearm, hand and fingers. More recently, [18] is developed specialized sensor fusion schema for stable estimation of fingertips with the data coming from LMC and sensorized glove sensors. [21] is introduced another kalman filter application with Myo Armband's inertial measurement unit to obtain better palm direction estimation at the measurement limits of LMC. Moreover, they used convolutional neural network classification to overcome drawbacks of Leap Motion's active finger distinguish on the measurement limits. Another kalman filter data fusion strategy proposed by [22], to achieve stable estimation of palm center with the position data gathered from LMC and velocity data gathered from Microsoft Kinect.

Finally, [23] introduced use of multi-LMCs to solve occlusion problem for teleoperative demonstration in the robotic system simulation. They analyzed different configurations of second LMC positioning to achieve optimal use of informations from the two sensors. However, in order to fuse data, they used the data from both LMCs to choose the more reliable one, not to obtain a smoother estimation. Moreover, their method is not applicable to real-time scenario and suffers from manual delay compensation for each acquisition.

III. Methodology

A. Self-calibration for coordinate transform

Goal of the proposed methodology is allowing arbitrary positioning of the LMC sensors, as seen in Fig. 1. However, LMC measures the hand frame with respect to LMC base reference frame where $\boldsymbol{{}^{L1}T_H}$ and $\boldsymbol{{}^{L2}T_H}$ represents measurements of hand frame with respect to the LMC-1 and LMC-2, respectively. These measurements can not be used for sensor fusion directly since reference frame is not identical. Hence, calibration matrix ($\boldsymbol{{}^{L1}T_{L2}}$) between reference frames is required. Only then, it is possible to estimate one sensor's measurements on the other sensor's reference frame. Accordingly, measurements acquired by LMC-1, could be estimated in the LMC-2 reference frame ($\boldsymbol{{}^{L2}\hat{T}_H}$) as follows:

$$\boldsymbol{{}^{L2}\hat{T}_H} = (\boldsymbol{{}^{L1}T_{L2}})^{-1}\ \boldsymbol{{}^{L1}T_H} \tag{1}$$

After the calibration, it is feasible to perform sensor fusion by using LMC-2 measurements ($\boldsymbol{{}^{L2}T_H}$) on it's own frame and LMC-1 measurements estimated in LMC-2 frame ($\boldsymbol{{}^{L2}\hat{T}_H}$).

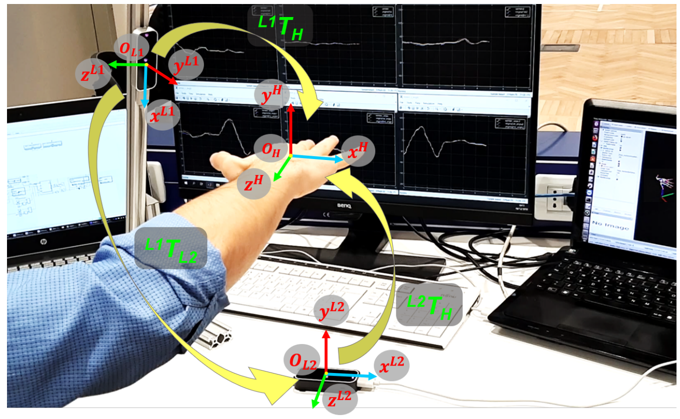


Fig. 1. Experimental setup and defined coordinate systems.

In consideration of arbitrary positioning of sensors for each experiment, calibration process designed to be performed automatically. In order to apply self-calibration algorithm, data is sequentially collected from two LMCs. To ensure the quality of calibration and a variety of workspace rules in Algorithm 1 are adopted:

Algorithm 1 Proposed Self-Calibration Algorithm

while $i < N_s$ **do**
  if $\|{}^{L1}z_{pos,k} - {}^{L1}z_{pos,k-1}\| > 1[mm]$ & $\|{}^{L2}z_{pos,k} - {}^{L2}z_{pos,k-1}\| > 1[mm]$ **then**
    if $C_{L1,k} > 0.8$ & $C_{L2,k} > 0.8$ **then**
      ${}^{L1}z_{buffer} \leftarrow {}^{L1}z_{pos,k}$
      ${}^{L2}z_{buffer} \leftarrow {}^{L2}z_{pos,k}$
    end if
  end if
end while
$\boldsymbol{{}^{L1}T_{L2}} \leftarrow Horn's\ Calibration({}^{L1}z_{buffer}, {}^{L2}z_{buffer})$

where ${}^{L1}z_{pos,k}$ and ${}^{L2}z_{pos,k}$ is the k-th instant palm position measurements with respect to the LMC-1 and LMC-2 respectively. $C_{L1,k}$ and $C_{L2,k}$ are the confidence level of the measurements acquired by LMC-1 and LMC-2. Finally, algorithm will stop when 500 samples (N_S) are collected by respecting these criteria. Note that, sampling frequency of sensors are $\approx 110\,[Hz]$ and data acquisition for calibration with described criteria usually completed in $30\,[sec]$ to $1\,[min]$. After the data acquisition, automatic calibration of sensor frames is achieved by using Horn's Method [24] between collected trajectories from LMC-1 and LMC-2. In order to overcome differences in linear measurements between sensors, calibration matrix computed as transformation matrix scaled in 3D. As a result, calibration matrix ($\boldsymbol{{}^{L1}T_{L2}}$) is obtained and LMC-1 measurements estimated in the LMC-2 reference frame ($\boldsymbol{{}^{L2}\hat{T}_H}$).

B. Sensor Fusion

1) Modeling: Hand model is treated as linear dynamical system in space and measurements considered as Gaussian distribution for modeling. Accordingly, state space model of hand based on discrete-time linear stochastic difference equation could be summed as [25]:

$$x_k = \Phi_k x_{k-1} + \Gamma_k u_k + w_k \tag{2}$$

$$z_k = H_k x_k + v_k \tag{3}$$

where $x_k \in \mathbb{R}^n$ is the state vector, $z_k \in \mathbb{R}^m$ is an measurement vector and $u_k \in \mathbb{R}^l$ is input vector. $\Phi_k \in \mathbb{R}^{nxn}$, $H_k \in \mathbb{R}^{mxn}$ and $\Gamma \in \mathbb{R}^{nxl}$ are state transition model, observation model and control input model. w_k and v_k are denotes the process and measurement noise respectively. They are assumed to be independent white gaussian noise (WGN):

$$w_K \sim WGN(0, Q) \tag{4}$$

$$v_k \sim WGN(0, R) \tag{5}$$

where Q is the covariance of process noise and R is the covariance of measurement noise. Also, note that for the case of hand motion analysis, control input u_k fired in the muscular system is considered as unknown. Additionally, hand can be modelled as rigid body in space. Linearized motion model for rigid body in space could be written as:

$$x_k = x_{k-1} + T_s \dot{x}_k + T_s^2/2\, \ddot{x}_k \tag{6}$$

$$\dot{x}_k = \dot{x}_{k-1} + T_s \ddot{x}_k \tag{7}$$

one that can model the acceleration term $\ddot{x}_k$, as a process noise. Therefore, hand motion model could be written as:

$$x_k = \Phi_k x_{k-1} + \Gamma_k w_k \tag{8}$$

Pointing out the difference, Γ_k is no longer control input, yet it represents the model that describes effect of process noise. In order to estimate pose of the hand, both position and orientation of hand frame is required. State for the hand position $(x_{pos,k})$ is denoted as:

$$x_{pos,k} = \begin{bmatrix} P_k & V_k \end{bmatrix}^T \tag{9}$$

where palm center position $P_k = \begin{pmatrix} P_{x,k} & P_{y,k} & P_{z,k} \end{pmatrix}$ and palm center velocity $V_k = \begin{pmatrix} V_{x,k} & V_{y,k} & V_{z,k} \end{pmatrix}$ in Cartesian coordinates of sensor reference frame. On the other hand, euler angles *(roll, pitch, yaw)* are adopted. State for hand orientation $(x_{orien,k})$ is defined as follows:

$$x_{orien,k} = \begin{bmatrix} \phi_k & \theta_k & \psi_k & \dot{\phi}_k & \dot{\theta}_k & \dot{\psi}_k \end{bmatrix}^T \tag{10}$$

Finally, by combining these assumptions, state transition and process noise effect model for the linearized motion of hand could be gathered as:

$$\Phi_k = \begin{bmatrix} 1 & 0 & 0 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 & T_s & 0 \\ 0 & 0 & 1 & 0 & 0 & T_s \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

$$\Gamma_k = \begin{bmatrix} T_s^2/2 & 0 & 0 \\ 0 & T_s^2/2 & 0 \\ 0 & 0 & T_s^2/2 \\ T_s & 0 & 0 \\ 0 & T_s & 0 \\ 0 & 0 & T_s \end{bmatrix} \tag{12}$$

Note that, same dynamical model (Φ_k, Γ_k) used for the both position and orientation estimation modeling. Observation model for position estimation (H_{pos}) includes position and velocity measurements of the palm:

$$H_{pos} = diag(1, 1, 1, 1, 1, 1) \tag{13}$$

On the other hand, orientation estimation only depends on the rotation matrix obtained from the hand frame. Observation model used in the orientation estimation (H_{orien}) is acquired by using this rotation matrix, and consequently *(roll, pitch, yaw)* angles are observed.

$$H_{orien} = diag(1, 1, 1, 0, 0, 0) \tag{14}$$

Rotation matrix of the hand frame constructed with the *palm normal* and *hand direction* measured from the LMC as it can be seen from the Fig. 2. Reference frame of hand (x^H, y^H, z^H) constructed as: $y^H = -palm\ normal$, $z^H = -hand\ direction$ and $x^H = palm\ normal \times hand\ direction$. In this line, rotation matrix from the *palm normal* and *hand direction* vectors measured by LMC is used to construct rotation matrix as the following equations:

$$cross\,product = palm\,normal \times hand\,direction \tag{15}$$

$$R(\alpha,\beta,\gamma) = \begin{bmatrix} -palm\,normal_x & -hand\,direction_x & cross\,product_x \\ -palm\,normal_y & -hand\,direction_y & cross\,product_y \\ -palm\,normal_z & -hand\,direction_z & cross\,product_z \end{bmatrix} \tag{16}$$

On the other hand, rotation matrix can be also represented by euler angles *(roll, pitch, yaw)*. Let us denote α as *roll* angle, β as *pitch* angle and γ as *yaw* angle. Rotation matrix could be represented for this notation as:

$$R(\alpha,\beta,\gamma) = R_z(\alpha)R_y(\beta)R_x(\gamma) = \\ \begin{bmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma \end{bmatrix} \tag{17}$$

In order to compute corresponding euler angles, method of [26] is used.

2) Kalman Filter: For the sensor fusion, kalman filter will be applied separately for both position and orientation estimation of the hand. Kalman filter is a recursive filter. In the each loop, it predicts the next state $\hat{x}_{k|k-1}$ and the covariance matrix $P_{k|k-1}$ as:

$$\hat{x}_{k|k-1} = \Phi_k \hat{x}_{k-1|k-1} \tag{18}$$

$$P_{k|k-1} = \Phi_k P_{k-1|k-1} \Phi_k^T + \Gamma_k Q_k \Gamma_k^T \tag{19}$$

Inside of the each loop, process is updated iteratively. Updated state estimate:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y_k} \tag{20}$$

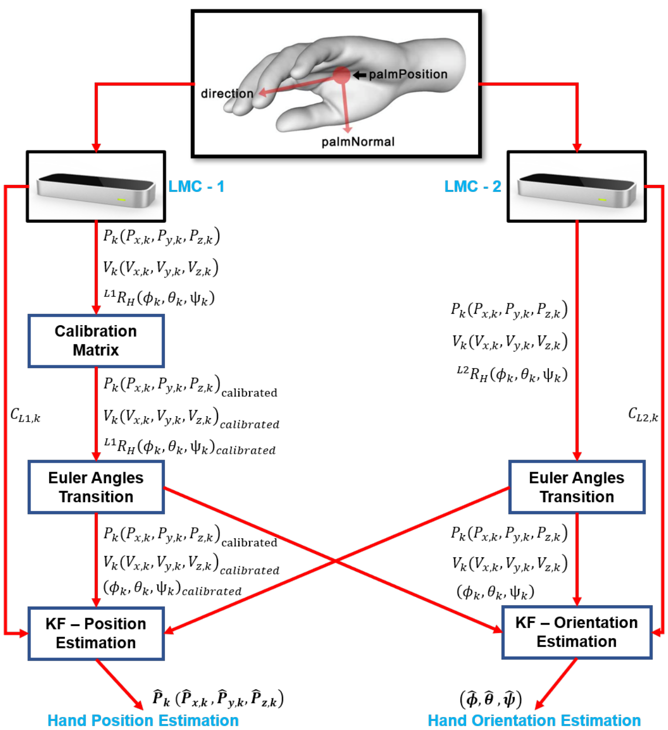


Fig. 2. Proposed adaptive sensor fusion schema.

where kalman gain(K_k) and measurement residual $(\tilde{y_k})$ defined as:

$$\tilde{y_k} = z_k - H_k \hat{x}_{k|k-1} \tag{21}$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \tag{22}$$

Residual variance matrix S_k is defined as:

$$S_k = H_k P_{k|k-1} H_k^T + R_k \tag{23}$$

Updated estimate of covariance matrix:

$$\hat{P}_{k|k} = (I - K_k H_k) P_{k|k-1} \tag{24}$$

$$\mathbb{E} = [\Gamma \, Q \, \Gamma^T] \tag{25}$$

$$Q = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{bmatrix} \tag{26}$$

$$\mathbb{E} = \begin{bmatrix} T_s^4/4 & 0 & 0 & T_s^3/2 & 0 & 0 \\ 0 & T_s^4/4 & 0 & 0 & T_s^3/2 & 0 \\ 0 & 0 & T_s^4/4 & 0 & 0 & T_s^3/2 \\ T_s^3/2 & 0 & 0 & T_s^4/4 & 0 & 0 \\ 0 & T_s^3/2 & 0 & 0 & T_s^4/4 & 0 \\ 0 & 0 & T_s^3/2 & 0 & 0 & T_s^4/4 \end{bmatrix} \tag{27}$$

In order to integrate multi leap motions with the proposed structure parallel kalman structure is used as discussed in [27]. Therefore, observation model (H), measurement noise (R) and instantaneous measurements (z_k) unified for parallel multi kalman structure as:

$$z_k = \begin{bmatrix} z_{k,L1} \\ z_{k,L2} \end{bmatrix} \tag{28}$$

$$H_k = \begin{bmatrix} H_{k,L1} \\ H_{k,L2} \end{bmatrix} \tag{29}$$

$$R = \begin{bmatrix} R_{L1} & 0 \\ 0 & R_{L2} \end{bmatrix} \tag{30}$$

3) Proposed Scheme for Adaptive Sensor Fusion: In the Fig. 2 proposed adaptive sensor fusion method with two separate kalman filters for hand position and orientation estimation is visualized. Measurement noise covariance for LMC-1 (R_{L1}) and LMC-2 (R_{L2}) are fixed at the beginning. These matrices, adaptively modified during the process with respect to the confidence level data incoming from the LMC-1 $C_{L1,k} \in \begin{bmatrix} 0 & 1 \end{bmatrix}$ and LMC-2 $C_{L2,k} \in \begin{bmatrix} 0 & 1 \end{bmatrix}$. Updated measurement noise covariance denoted as $R^*_{L1,k}$ and $R^*_{L2,k}$ for LMC-1 and LMC-2 respectively. Updating rule is given as follows:

$$R^*_{L1,k} = R_{L1}/(C_{L1,k} + \alpha) \tag{31}$$

$$R^*_{L2,k} = R_{L2}/(C_{L2,k} + \alpha) \tag{32}$$

where α is arbitrarily small value that prevents resulting infinite numbers when confidence level reaches to zero.

C. System Description

In line with the proposed algorithm, multi-sensor processing and data analysis done with efficient, communicating multi-computers, as shown in Fig. 3. ROS[3] network used to transmit data between computers. Synchronous data transfer ensured by using timestamps at the ROS messages.

The first computer has an i7-4720HQ CPU 2.60 GHz processor and 8 GB RAM, and collects hand frame data from LMC-1 and the second computer has an i7-7700HQ 2.8 GHz CPU, 8 GB GeForce 1070 GPU and 16 GB RAM, and gathers hand frame data from LMC-2. The sampling rate is set at 110Hz for both devices.

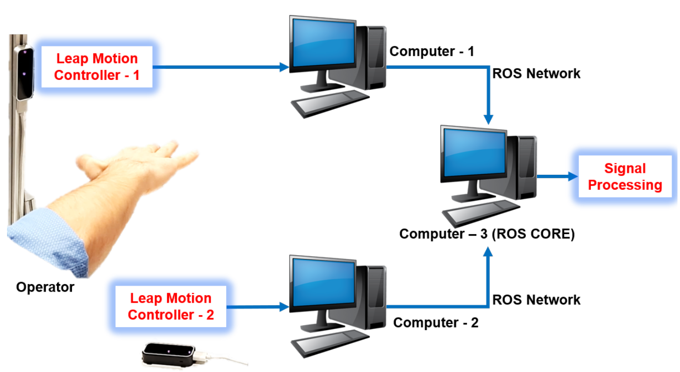


Fig. 3. Overview of system description.

Finally, sensor fusion performed by the third computer with an i9-9900K 3.6 GHz CPU, 8 GB Quadro M5000 GPU and 64 GB of RAM. The sensors of the proposed hand gesture recognition system are listed as follows:

A Leap Motion Controller (Leap Motion, California, United States), which consists of two cameras and three infrared LEDs, tracking infrared light with a wavelength of 850 nanometers (up to 115 Hz);

[3]Robot Operating System, http://www.ros.org/

IV. Results

Results are investigated in three sections. Firstly, results of reference frame registration between local reference frames of LMCs are given in order to show calibration performance. In the following sections, sensor fusion results are compared to the raw data gathered from LMCs. In the second section of results, hand motion is tracked without occlusion and results showing the smoothness is given. Lastly, occlusion case where one LMC lost detection as seen in Fig. 4 is investigated and results showed that proposed methodology is able to track during occlusion occurs on one LMC.

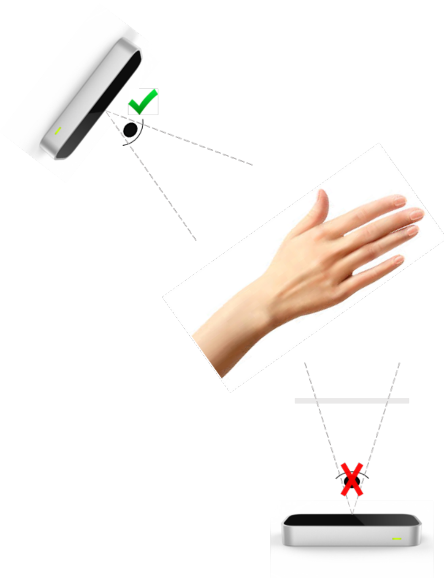


Fig. 4. When the case of occlusion occurs on one LMC and therefore only LMC detects

A. Calibration

Resulting calibration is given in Fig. 5. On the left figure, measurements acquired in the own reference frame of LMCs before the calibration step is represented. As it can be seen from the figure, it is not possible to use sensor fusion without calibration. On the right figure, measurements taken by the LMC-1, calibrated to the LMC-2 reference frame by the Algorithm 1 defined in the previous sections. For a validation of calibration result, mean absolute error (MAE) is computed as:

$$MAE = \frac{1}{N}\sum_{k=1}^{N} \| P_{L1_{calibrated},k} - P_{L2,k} \| \tag{33}$$

As a result of Horn's calibration method, MAE achieved as $MAE = 5.4728[mm]$ between the calibrated LMC-1 estimations and LMC-2 measurements.

B. Sensor Fusion: No Occlusion

In this first scenario, two leap motions could detect the hand without occlusion during acquisitions, sensor fusion estimates hand pose smoothly thanks to the proposed methodology

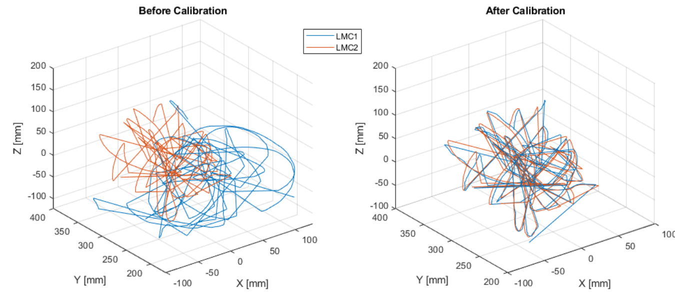


Fig. 5. At the left, raw measurements from LMCs plotted in their own reference frame before calibration and at the right, calibrated LMC-1 estimations and LMC-2 measurements are visualized as both in the reference frame of LMC-2

with adaptive measurement covariance updates. In the Fig. 6 resulting sensor fusion data with the raw data acquired from LMCs are given for position and orientation of the palm $\begin{pmatrix} P_x & P_y & P_z & \phi & \theta & \psi \end{pmatrix}$.

As it can be seen from the graph, sensor fusion provides smoother trajectory than the raw data. Especially for the orientation estimation, noisy measurements fused smoothly without missing major information. In order to validate smoothness of the proposed method, jerk of the palm position $[mm/s^3]$ and palm orientation $[rad/s^3]$ is computed as it can be seen in the Table I.

TABLE I
RMS of Jerk Comparison Between Raw Sensors Data (L1, L2) and Sensor Fusion Algorithm (SF)

Sensor	RMS of Jerk for Output States $[mm/s^3]$					
	$P_{x,k}$	$P_{y,k}$	$P_{z,k}$	ϕ_k	θ_k	ψ_k
L1	1.595	1.871	1.674	0.682	0.824	1.297
L2	1.460	1.753	1.346	0.730	0.504	1.208
SF	0.042	0.042	0.043	0.121	0.104	0.215

C. Sensor Fusion: Occlusion on One LMC

The most importantly, proposed adaptive sensor fusion algorithm provides continuous estimations, even one sensor is not able to detect the hand Fig. 4 due to the occlusion. In this case, confidence level incoming from the sensor, which is not able to detect the hand, will be equal to zero. According the proposed methodology, this will lead very big measurement noise covariance for that specific sensor. In this way, continuous tracking is feasible even occlusion occurs on the one LMC. Results are given in Fig. 7 for $\begin{pmatrix} P_x & P_y & P_z & \phi & \theta & \psi \end{pmatrix}$. Regions with straight line indicate that sensor was not able to detect hand. During the experiments, this situation simulated by manually placing an flat object in front of the sensor.

Overall, proposed adaptive sensor fusion methodology was able to track hands continuously, and showed smooth transitions between states that two sensors tracking and only one sensor tracking while one sensor were not able to track.

V. Conclusion

In conclusion, adaptive sensor fusion algorithm proposed for hand pose tracking via multi-LMCs. Thanks to the adaptive

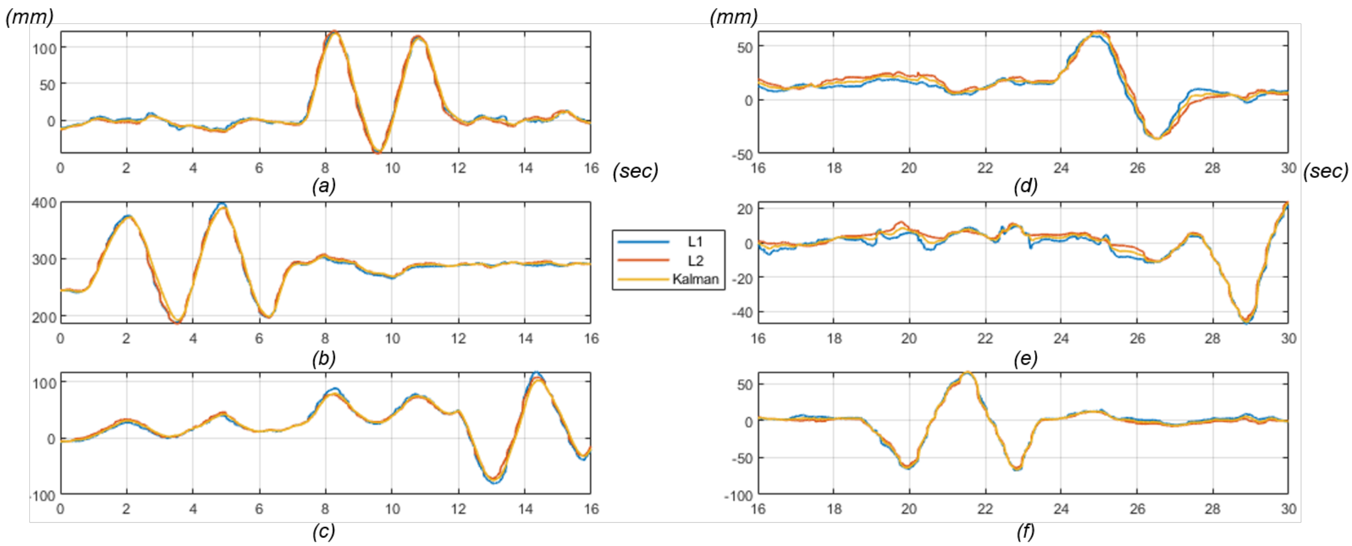


Fig. 6. Estimations when two of the LMC tracks hand

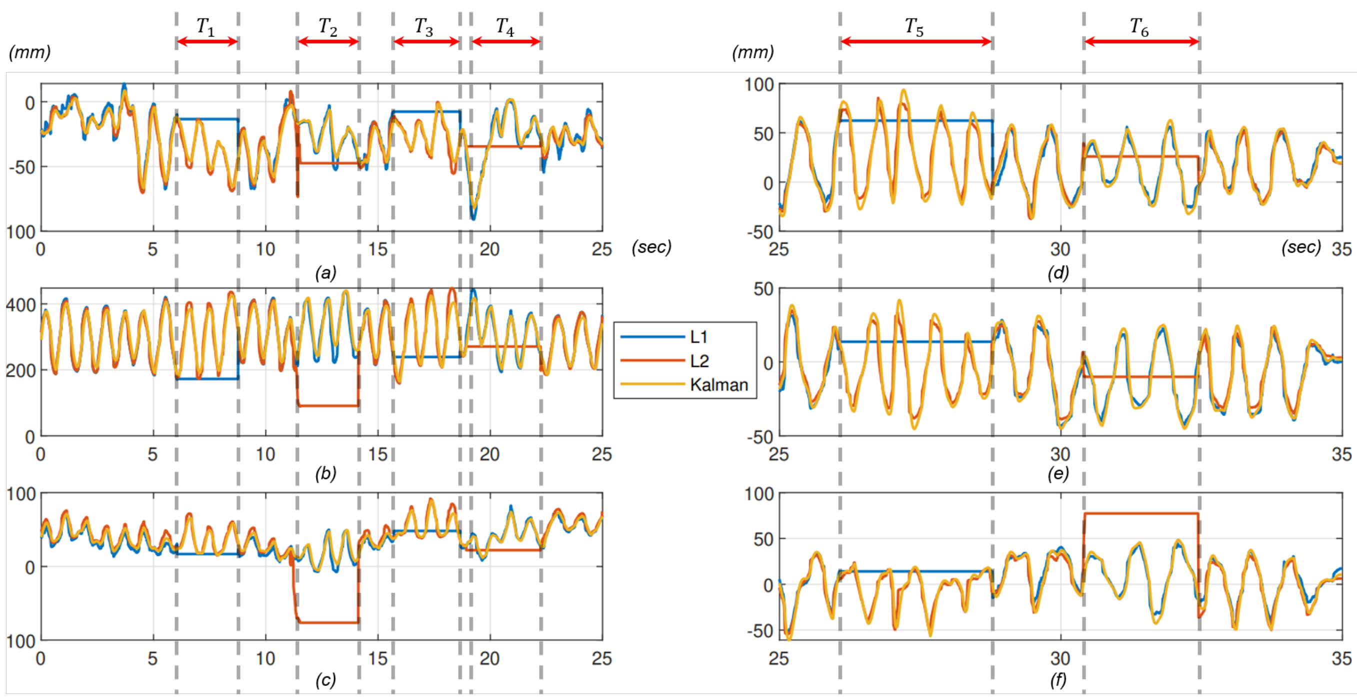


Fig. 7. Estimations when one LMC lost tracking

measurement covariance update, continuous and stable pose estimation achieved despite the impermanent detection of hands. Proposed algorithm is able to work in real-time. As a future works, real-time applications such as teleoperating robots could be considered for further validation. In order to avoid from Gimbal Lock singularity occurs naturally from the euler angles, quaternions could be used for orientation estimation of the hand.

REFERENCES

[1] R. de la Barré, P. Chojecki, U. Leiner, L. Mühlbach, and D. Ruschin, "Touchless interaction-novel chances and challenges," in *International Conference on Human-Computer Interaction*. Springer, 2009, pp. 161–169.

[2] G. J. Kim, *Human–Computer Interaction: Fundamentals and Practice*. Auerbach Publications, 2015.

[3] R. Lun and W. Zhao, "A survey of applications and human motion recognition with microsoft kinect," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, no. 05, p. 1555008, 2015.

[4] X. Chang, Z. Ma, M. Lin, Y. Yang, and A. G. Hauptmann, "Feature interaction augmented sparse learning for fast kinect motion detection," *IEEE transactions on image processing*, vol. 26, no. 8, pp. 3911–3920, 2017.

[5] W. Lu, Z. Tong, and J. Chu, "Dynamic hand gesture recognition with leap motion controller," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1188–1192, 2016.

[6] Y. Pulijala, M. Ma, and A. Ayoub, *VR Surgery: Interactive Virtual Reality Application for Training Oral and Maxillofacial Surgeons using Oculus Rift and Leap Motion*. Cham: Springer International Publishing, 2017, pp. 187–202.

[7] D. Bassily, C. Georgoulas, J. Guettler, T. Linner, and T. Bock, "Intuitive and adaptive robotic arm manipulation using the leap motion controller," in *ISR/Robotik 2014; 41st International Symposium on Robotics*, June 2014, pp. 1–7.

[8] Cao Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using microsoft kinect," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 44–52.

[9] F. Despinoy, N. Zemiti, G. Forestier, A. Sánchez, P. Jannin, and P. Poignet, "Evaluation of contactless human–machine interface for robotic surgical training," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 1, pp. 13–24, Jan 2018. [Online]. Available: https://doi.org/10.1007/s11548-017-1666-6

[10] M. Kim, C. Jeon, and J. Kim, "A study on immersion and presence of a portable hand haptic system for immersive virtual reality," *Sensors*, vol. 17, no. 5, p. 1141, 2017.

[11] A. Mewes, B. Hensen, F. Wacker, and C. Hansen, "Touchless interaction with software in interventional radiology and surgery: a systematic literature review," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 2, pp. 291–305, Feb 2017. [Online]. Available: https://doi.org/10.1007/s11548-016-1480-6

[12] P. Julien, L. Tommaso, A. Lounis, M. Benassarou, P. Mathieu, D. Bernot, and S. Aubry, "Leap motion gesture control with carestream software in the operating room to control imaging," *Surgical innovation*, vol. 22, 05 2015.

[13] H. Su, S. E. Ovur, X. Zhou, W. Qi, G. Ferrigno, and E. D. Momi, "Depth vision guided hand gesture recognition using electromyographic signals," *Advanced Robotics*, vol. 0, no. 0, pp. 1–13, 2020. [Online]. Available: https://doi.org/10.1080/01691864.2020.1713886

[14] S. E. Ovur, F. Candan, A. Beke, and T. Kumbasar, "Yaft: A fuzzy logic based real time two-wheeled inverted pendulum robot," in *2018 6th International Conference on Control Engineering Information Technology (CEIT)*, Oct 2018, pp. 1–6.

[15] B. Penelle and O. Debeir, "Multi-sensor data fusion for hand tracking using kinect and leap motion," in *Proceedings of the 2014 Virtual Reality International Conference*, 2014, pp. 1–7.

[16] E. C. Silva, E. W. Clua, and A. A. Montenegro, "Sensor data fusion for full arm tracking using myo armband and leap motion," in *2015 14th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*. IEEE, 2015, pp. 128–134.

[17] M. Mohandes, S. Aliyu, and M. Deriche, "Prototype arabic sign language recognition using multi-sensor data fusion of two leap motion controllers," in *2015 IEEE 12th International Multi-Conference on Systems, Signals & Devices (SSD15)*. IEEE, 2015, pp. 1–6.

[18] G. Ponraj and H. Ren, "Sensor fusion of leap motion controller and flex sensors using kalman filter for human finger tracking," *IEEE Sensors Journal*, vol. 18, no. 5, pp. 2042–2049, March 2018.

[19] G. Du and P. Zhang, "A markerless human–robot interface using particle filter and kalman filter for dual robots," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2257–2264, April 2015.

[20] G. Du, P. Zhang, and X. Liu, "Markerless human–manipulator interface using leap motion with interval kalman filter and improved particle filter," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 694–704, April 2016.

[21] J. Chen, C. Liu, R. Cui, and C. Yang, "Hand tracking accuracy enhancement by data fusion using leap motion and myo," in *International Conference on Unmanned Systems and Artificial Intelligence (ICUSAI)*. IEEE, 2019.

[22] C. Li, A. Fahmy, and J. Sienz, "An augmented reality based human-robot interaction interface using kalman filter sensor fusion," *Sensors*, vol. 19, no. 20, p. 4586, 2019.

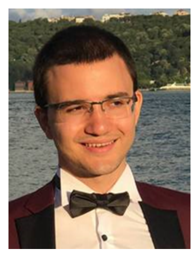
[23] H. Jin, Q. Chen, Z. Chen, Y. Hu, and J. Zhang, "Multi-leapmotion sensor based demonstration for robotic refine tabletop object manipulation task," *CAAI Trans. Intell. Technol.*, vol. 1, pp. 104–113, 2016.

[24] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society A*, vol. 4, pp. 629–642, 04 1987.

[25] M. Gerdin, "Identification and estimation for models described by differential-algebraic equations," Ph.D. dissertation, Institutionen för systemteknik, 2006.

[26] G. G. Slabaugh, "Computing euler angles from a rotation matrix," *Retrieved on August*, vol. 6, no. 2000, pp. 39–63, 1999.

[27] D. Willner, C. Chang, and K. Dunn, "Kalman filter algorithms for a multi-sensor system," in *1976 IEEE Conference on Decision and Control including the 15th Symposium on Adaptive Processes*. IEEE, 1976, pp. 570–574.



SALIH ERTUG OVUR received his B.Sc. degree in Mechanical Engineering in Istanbul Technical University (Turkey), in 2017 with 3rd best student award. Also he received secondary B.Sc. degree from Control and Automation Engineering in Istanbul Technical University (Turkey), in 2018 with 1st best student award. He is currently a M.Sc. student in Automation and Control Engineering at Politecnico di Milano (Italy). He is working in the field of surgical robotics, in particular, his research focuses on developing assistive technologies for surgeons.

HANG SU received the M.Sc. degree in control theory and control engineering in South China University of Technology, Guangzhou, China, in 2015. He obtained his PhD Degree in 2019 as a member of the Medical and Robotic Surgery group (NEARLab) in Politecnico di Milano, Milano, Italy. He participated in the EU funded project (SMARTsurg) in the field of Surgical Robotics. Dr. Hang Su is currently a Research Fellow in the Department of Electronics, Information and Bioengineering (DEIB) of Politecnico Di Milano. He has served as a reviewer for over 30 scientific journals, like IEEE Transaction on Biomedical Engineering, IEEE/ASME Transactions on Mechatronics, IEEE Transaction on Automation and Engineering, IEEE Transaction on Cybernetics, IEEE Transactions on Systems, Man, and Cybernetics: Systems, etc.. He is currently Special Session Chair of IEEE International Conference on Advanced Robotics and Mechatronics (ICARM 2020).

He has published several papers in international conferences and journals and has been awarded ICRA 2019 travel grant. His main research interests include control and instrumentation in medical robotics, human-robot interaction, surgical robotics, deep learning, bilateral teleoperation, etc..

WEN QI (S'19) received the M.Sc. degree in Control engineering from the South China University of Technology, Guangzhou, China, in 2015. Her first-authored paper is awarded the finalist of T J Tarn's Best Paper Award on Control Applications on IEEE WCICA 2014. Now she is pursuing a Ph.D. Degree as a member of the Laboratory of Biomedical Technologies (TBMLab) in Politecnico di Milano, Milano, Italy. Her main research interests include machine learning, deep learning and signal processing algorithms in wearable medical devices.

ELENA DE MOMI (M'07-SM'18) received the M.Sc. and Ph.D. degrees in biomedical engineering from the Politecnico di Milano, Milan, Italy, in 2002 and 2006, respectively.

She is an Associate Professor with the Department of Electronics, Information and Bioengineering, Politecnico di Milano. She is co-founder of the Neuroengineering and Medical Robotics Laboratory, in 2008, being responsible of the Medical Robotics section. IEEE Senior Member, she is currently Associate Editor of the Journal of Medical Robotics Research, of the International Journal of Advanced Robotic Systems, Frontiers in Robotics and AI and Medical & Biological Engineering & Computing. From 2016 she has been an Associated Editor of IEEE ICRA, IROS and BioRob and she is currently Publication Co-Chair of ICRA 2019. She is responsible for the lab course in Medical Robotics and of the course on Clinical Technology Assessment of the MSc degree in Biom. Eng. at Politecnico di Milano and she serves in the board committee of the PhD course in Bioengineering.

GIANCARLO FERRIGNO received the M.Sc. degree in electrical engineering and the Ph.D. degree in bioengineering from the Politecnico di Milano, Milan, Italy.

He is the Founder of the Neuroengineering and Medical Robotics Laboratory with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, in 2008, and a Lecturer of Medical Robotics. He is a Full Professor with the Politecnico di Milano. He has been the European Coordinator of three FP7 EU projects on ICT. Two of them, ROBOCAST (STREP 2008-2010) and ACTIVE (Integrated project 2011-2015) are in the field of the Surgical Robotics. MUNDUS (STREP 2010-2013) is in the field of Assistive and Rehabilitative Robotics. He has co-authored 20 papers (ISI Web of Knowledge) in the robotic field from 2011 to 2014. He is working in the JWG9 ISO standard group for Surgical Robots collateral standard and organized several workshops in the surgical robotics for the last three years.