Hand Pose Detection in HMD Environments by Sensor Fusion using Multi-layer Perceptron

Luc Cong Vu, Bum-Jae You
Dept. of HCI & Robotics Engineering, Univ. of Science and Technology (UST)
Center of Human-centered Interaction for Coexistence
Seoul, South Korea
vucongluc2018@gmail.com, ybj@kist.re.kr

Abstract— **The paper proposes a sensor fusion method to detect the pose of user's hand in head-mount display environments by using a Leap Motion Camera (LMC) with simple circular artificial markers on surfaces of a box wearing on the back of hand and two IMU sensors. One IMU sensor is located on the box and the other IMU sensor is fixed with the camera. Multi-layer Perceptron (MLP) is adopted to transform the hand's pose in IMU coordinate system into the pose in global coordinate system attached at LMC by minimizing Mean Square Error (MSE) for Virtual Reality (VR)/Mixed Reality (MR) applications. The pose detection results are compared with poses of bare hand captured by LMC while the estimated data after transformation is fitted well with reference data in the sense that the average of mean difference for each roll, pitch and yaw angle is around 3.54 degree. It is applied successfully to track and estimate the pose of user's hand in around 70Hz.**

Keywords- **Multi-layer Perceptron, Hand pose detection, Sensor Fusion, IMU sensors, Leap Motion camera, Human Interaction.**

I. INTRODUCTION

Recently, Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) become popular and start having the impact on human life, such as in education, working, communication, gaming, entertainment and etc. Many companies spend a large amount of resources and focus on researches to develop VR/AR technologies. One of the most important things is to develop technologies in state-of-art for hand tracking to further enhance VR/AR experiences. Hand pose detection is critical in HMD-based virtual reality and/or augmented reality environments for natural interaction with virtual information using user's hand. There have been proposed a few sensors for hand pose tracking such as HTC Vive tracker [1], Kinect camera [2], and Leap Motion Camera (LMC) [3]. LMC has much interests since it has a potential to provide an interaction by bare hand with small size and more flexibility than others. There are several approaches using LMC to show a demonstration on hand interaction with virtual information [4-8]. However, LMC has the low accuracy for fingertip detection, especially, the movement of thumb finger.

The smart gloves or wearable devices are investigated to enhance the performance of finger tracking such as Cyber Glove III [9], Perception Neuron glove [10] and Wireless Data glove [11]. The paper [12] presents the wearable glove using inertial and magnetic sensors for hand tracking. The sensor system can estimate the joints' values of the hand as well as the hand rotation. However, these devices have not provided the pose of user's hand in HMD's global coordinate system. So it is hard to use for VR/MR applications that we should know the orientation in global coordinate. Sensor fusion methods have been proposed to overcome previous limitations. In [13], a sensor fusion approach by the combination of an IMU sensor with a fixed external camera is proposed for precise position estimation of fiducial markers attached with an IMU on the back of hand. Its 3D pose is decided from the IMU sensor data and compensated by position data from the video tracking using extended Kalman filter model. [14] The hand tracking glove is designed by the flex sensor to track the finger motion while marker and IMU sensor are attached to the back of the glove to detect the pose of glove. In these approaches, however, the pose of a target (hand) cannot be detected when one or more markers are lost during changes of line-of-sights by motions of user's hand and HMD. To overcome the difficulty, there is under development a new approach to detect hand pose by using simple circular artificial markers on a box with an IMU sensor attached at the back of hand tracked by LMC with other IMU sensor.

In this paper, a sensor fusion method to detect the pose of user's hand in HMD environments by using a LMC with simple circular artificial markers on surfaces of a box wearing on the back of hand and two IMU sensors. One IMU sensor is located on the box and the other IMU sensor is fixed with the camera. Hand's orientation is tracked by relative Euler angles between two IMU sensors. And, Multi-layer Perceptron (MLP) is adopted to find a transformation from the hand's pose in IMU coordinate system into the pose in global coordinate system attached at LMC by minimizing mean square errors for VR/MR applications. It is experimented successfully in real-time, 70Hz, under popular PC environments for virtual reality.

The rest of the paper is organized as follows. Section II introduces the proposed sensor system while section III describes the proposed pose detection algorithms. Section IV shows the experimental results and conclusions and future research are drawn in Section V.

II. PROPOSED SENSOR SYSTEM

In order to capture finger joints movement, hand motion capture devices such as data gloves and/or wearable devices have been developed. For example, a wearable exoskeleton device, **'CHICAP'** was introduced recently for accurate measurement of finger motions as shown in Fig. 1. [15] The device captures accurately motions of three fingers including thumb, index finger and middle finger.

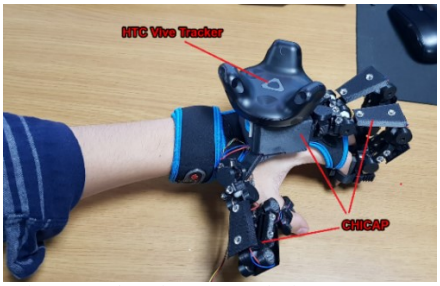


Fig. 1. Hand motion capture device 'CHICAP'

For natural spatial interaction of user's hand with virtual objects, however, pose information – 3D position and orientation – of user's hands in HMD's global coordinate system is needed more. Even though HTC Vive trackers - attached on the back of hand as shown in Fig. 1 - are used widely to track the pose of user's hand, it is really big and heavy for users, and it can be used only for HTC Vive lighthouse system. So, there is required a new sensor system and hand pose detection approach that is applicable for other HMD systems except for HTC Vive HMD and is useful for mobile environments.

A sensor system is proposed by using a LMC, two IMU sensors, and a box with circular markers as shown in Fig. 2. The LMC is a small-size IR stereo camera from Leap Motion Inc. whose frame rate is 120 Hz and resolution is 2 x 640 x 480. The IMU sensor is a MPU-9250 from Invensense Inc. providing 9-DOF motion data combining 3-axis gyroscope, 3-axis accelerometer and 3-axis magnetometer. A LMC is attached in front of a HMD. An IMU sensor is fixed together with the LMC in order to track orientation of the HMD when users rotate their head. A box with two circular markers in each surface is located at the back of user's hand. One IMU sensor is also attached on the box in order to track hand orientation.

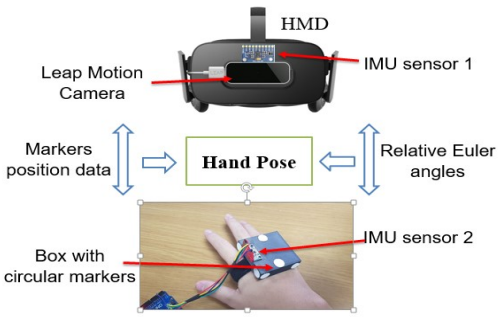


Figure 2. Sensor System Overview

Kalman filter [16] is used to fuse the data of three sensors in an IMU sensor and to remove noise from accelerometer and drift from gyroscope. The output data of Kalman Filter is Euler angles - Roll, Pitch, and Yaw corresponding to the rotation angles in x-axis, y-axis and z-axis, respectively.

III. PROPOSED POSE DETECTION ALGORITHMS

A. *Surface Classification and Position Detection*

Since each surface of the box has the same circular makers, the markers on one surface may be confused with markers on the other surface when users rotate his/her hand if visual information is used only. So, the position of the box (i.e. user's hand) is decided by using relative relationship between positions of makers and the center of the box by classifying a surface that includes two circular markers tracked by LMC. There was developed a MLP-based marker classification approach whose inputs are relative Euler angles between two IMU sensors and positions of markers as shown in Fig 3. [17]

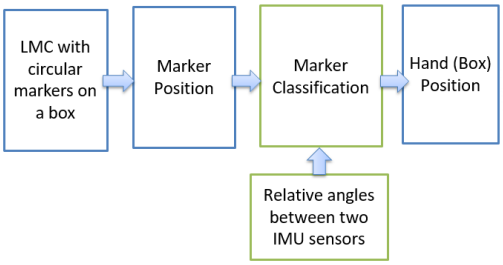


Figure 3. Hand Position Detection

B. *MLP-based Orientation Detection*

For real-time hand interaction with virtual objects in HMD environments, the relative orientation between two IMU sensors has to be transformed into the orientation in the reference coordinate of the LMC. MLP is adopted to find the transformation ($\hat{T}$) between the reference coordinates of the IMU sensor and the LMC. Then, it is applied for hand orientation transformation as 'Coordinate Transformation' block in Fig. 4. The hand orientation with respect to a global coordinate attached at LMC and/or HMD is decided finally by the proposed algorithm shown in Fig. 4.

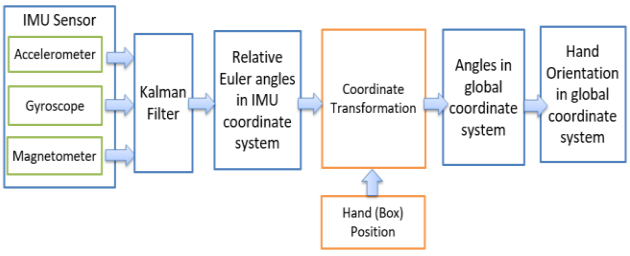


Figure 4. Hand orientation detection

Firstly, a sensor system in Fig. 5 is designed to collect data for training the MLP model. The box is worn at the palm of hand since the LMC can track only the bare hand and determine the pose of hand. The relative orientation

($\text{Roll}_{I_1^2}$, $\text{Pitch}_{I_1^2}$, $\text{Yaw}_{I_1^2}$) between two IMU sensors are collected and used as input data of the MLP model. Hand position in x-axis and y- axis directions is also captured and used for input data of the MLP model since the pose of hand is changed even in case of only translating the camera position while the user's hand has not moved or rotated.

So, five elements ($\text{Roll}_{I_1^2}$, $\text{Pitch}_{I_1^2}$, $\text{Yaw}_{I_1^2}, x, y$) are used as input data of the MLP model. The hand orientation, (Roll_R, Pitch_R, Yaw_R), detected by the LMC is used as output data of the MLP for training.

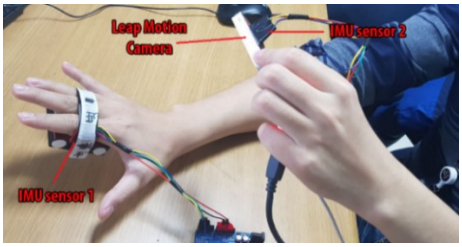


Figure 5. Sensor system for data collection

Secondly, by moving and rotating the hand in the field of view of the LMC, two data sets are captured and shown in Fig. 6. The left one is the 3D graph for relative Euler angles between two IMU sensors (blue color) and angles captured by the LMC with bare hand (green color). Each axis shows roll, pitch, and yaw axis, respectively. The right one is the line graph for hand position in x-axis and y- axis directions. Blue line presents x-coordinate while the y-coordinate is shown by green line.

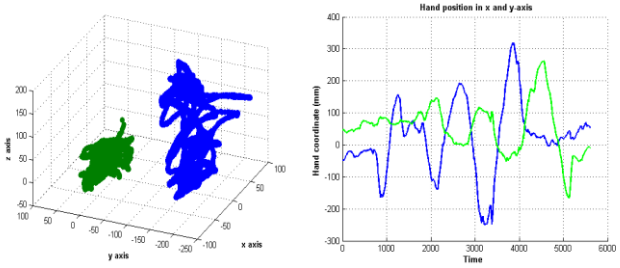


Figure 6. Collected data graph for training model

After collecting data for training, the MLP model is built as shown in Fig. 7. Input layer includes five elements ($\text{Roll}_{I_1^2}$, $\text{Pitch}_{I_1^2}$, $\text{Yaw}_{I_1^2}, x, y$) while (Roll_R, Pitch_R, Yaw_R) capturing from the LMC with a bare hand is used as data of output layer for training the MLP model. Hidden layer is trained by minimizing Mean Square Error (MSE) in Eq. (1) below. The model is represented by each couple weight matrices and bias vectors (W_i, b_i) between each layer.

$$MSE = \frac{1}{n}\sum_{i=0}^{n} D_i \qquad (1)$$

In which, D_i is the squared error of sample i^{th} between estimated data after transforming $\hat{T}$ into global coordinate, (Roll_L, Pitch_L, Yaw_L) and the reference data captured by the LMC with bare hand, (Roll_R, Pitch_R, Yaw_R).

$$D_i = \sqrt{\begin{matrix}(Roll_L^{(i)} - Roll_R^{(i)})^2 + (Pitch_L^{(i)} - Pitch_R^{(i)})^2 \\ +(Yaw_L^{(i)} - Yaw_R^{(i)})^2\end{matrix}} \qquad (2)$$

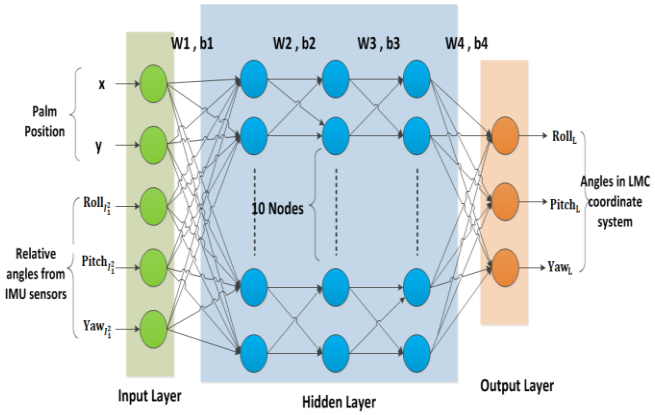


Figure 7. MLP model for transformation with 3 hidden layers and 10 nodes in each hidden layer.

To evaluate the model, it is assumed that if $D_i < \varepsilon$ then the estimated data is approximately same with output data. And the pose in IMU coordinate frame is transformed successfully into the pose in global coordinate frame. The accuracy of model is defined by Eq. (3). The threshold ε is selected experimentally.

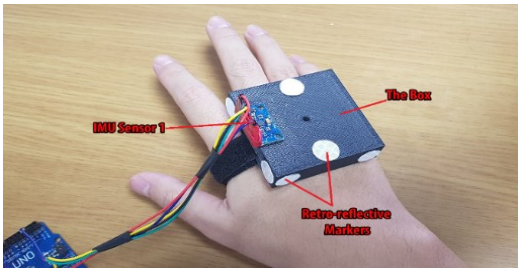
$$Accuracy = \frac{Number\ of\ samples\ (D_i < \varepsilon)}{Number\ of\ samples} * 100\% \qquad (3)$$

Finally, the pose of user's hand is determined by combining the position and orientation in Fig. 3 and Fig. 4.

IV. Experimental Results

A. Experimental Setup

The implemented sensors are shown in Fig. 8. The first one is the box with two retro-reflective markers on each surface and an IMU sensor to track hand orientation. The second is the LMC with a fixed IMU sensor that will be attached at a HMD.



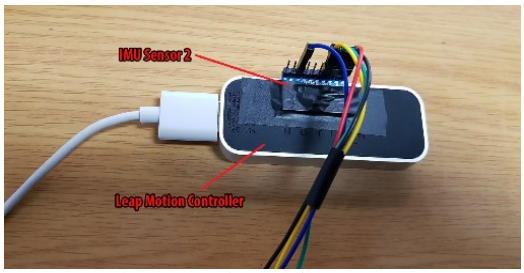


Figure 8. Implemented Sensors

B. *Training the MLP for transformation*

The collected data by the sensor system in Fig. 5 is captured around 20000 samples and separated randomly and independently into two parts. The first part with 80% sample data is used for training the MLP model while the other with 20% sample data is used for testing and calculating the accuracy. The number of hidden layers and units in each hidden layer is changed in training step. The accuracy and MSE are calculated with the threshold $\varepsilon = 10$ in Eq. (3). The results are shown on Table I.

TABLE I. THE RESULTS AFTER TRAINING WITH $\varepsilon = 10$

The number of hidden layers	The number of nodes in each hidden layer	Accuracy (%)	MSE (Degree)
1 layer	5 nodes/ 1 layer	69.72%	9.57
1 layer	10 nodes/ 1 layer	76.54%	8.14
2 layers	10 nodes/ 1 layer	84.61%	6.55
3 layers	10 nodes/ 1 layer	91.13%	5.21
3 layers	20 nodes/ 1 layer	98.93%	3.25
3 layers	50 nodes/ 1 layer	99.78%	1.82

As the results, if the complexity of model is increased then the accuracy is also increased while MSE is decreased. However, if the large number of hidden layers is used then it is hard to implement on a project in visual C++ environment. So, the model with 3 hidden layers and 10 nodes for each layer is chosen and drawn as shown in Fig. 7. The accuracy of model is around 91% while MSE is around 5.21.

C. *Hand Pose Detection*

After training the MLP model, it is applied to transform the orientation from IMU sensor coordinate frame into LMC global coordinate frame in real time.

The coordinate system in different colors is drawn for visualization testing of hand's pose. The Fig. 9 shows some poses of user's hand. The detected coordinate frame is drawn in different color – blue, green and red color corresponding with x-axis, y-axis and z-axis, respectively. The position of original coordinate system is at the center point of the box. The y-axis is shown in the direction of hand while z-axis is drawn in the direction of the normal vector of palm.

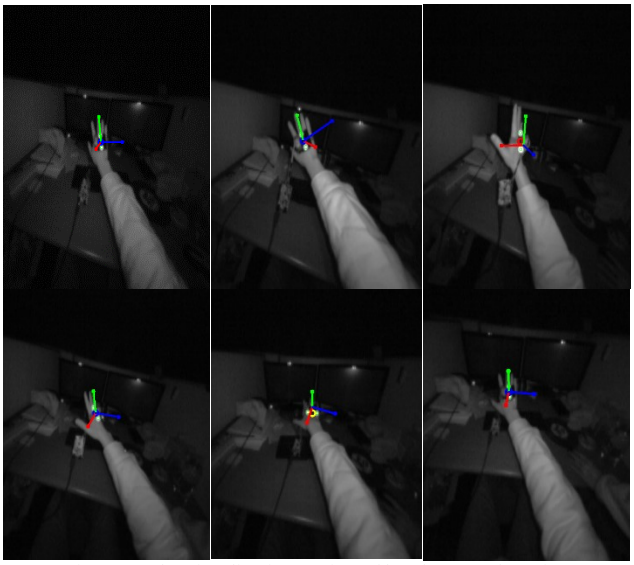


Figure 9. The visualization testing of hand's pose detection

As the results, correct directions are shown when users move and rotate the hand in real time.

D. *Hand Pose Evaluation*

In order to evaluate the accuracy of hand orientation detection using a MLP model to transform the coordinate system, the bare hand orientation by LMC is used as the reference orientation. The experimental system is designed as shown in Fig. 5. One IMU sensor is attached on the box while the other is attached on the LMC. The box is worn at the palm of hand since LMC can track only the bare hand and determine the pose of hand as the reference pose.

The pose of user's hand by two IMU sensors is transformed into global coordinate system using the trained MLP model. The transformed hand orientation is compared with the reference pose of the bare hand by LMC. Two coordinate frames are drawn in Fig. 10 for visual comparison. The green color coordinate system presents the pose from IMU sensors after transformation while the red color coordinate system shows the reference pose of bare hand tracking by LMC. As the results, two coordinate frames are fitted together in three directions.

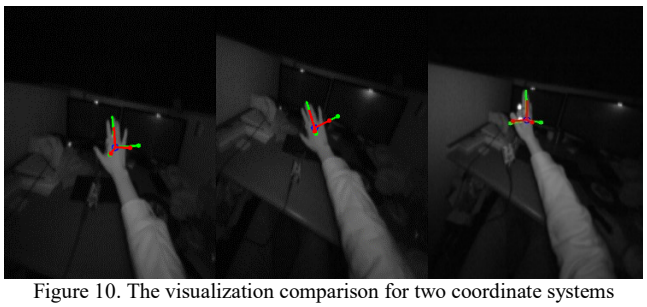


Figure 10. The visualization comparison for two coordinate systems

The hand's pose representing by Euler angles is captured in real time when the hand is moving in space freely and shown in 3D graph in Fig. 11.

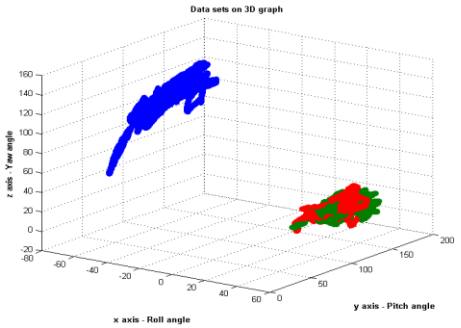


Figure 11. 3D graph for data

Each axis shows roll, pitch, and yaw data, respectively. The blue points present the relative Euler angles between two IMU sensors. Green points draw the angles of bare hand as the reference pose captured by LMC while red points show the estimated angles after applying the coordinate transformation. From the 3D graph, it is expected the hand's pose in IMU coordinate frame is transformed successfully into global coordinate frame.

To analyze the accuracy in quantification, the mean of difference between estimated angles that is transformed ($\hat{T}$)into global coordinate frame (Roll_L, Pitch_L, Yaw_L) and reference angles (Roll_R, Pitch_R, Yaw_R) is calculated by Eq. (4).

$$M_{Roll} = \frac{1}{n}\sum_{i=1}^{n}\left|d_{Roll}^{(i)}\right| = \frac{1}{n}\sum_{i=1}^{n}\left|Roll_L^{(i)} - Roll_R^{(i)}\right|$$
$$M_{Pitch} = \frac{1}{n}\sum_{i=1}^{n}\left|d_{Pitch}^{(i)}\right| = \frac{1}{n}\sum_{i=1}^{n}\left|Pitch_L^{(i)} - Pitch_R^{(i)}\right| \quad (4)$$
$$M_{Yaw} = \frac{1}{n}\sum_{i=1}^{n}\left|d_{Yaw}^{(i)}\right| = \frac{1}{n}\sum_{i=1}^{n}\left|Yaw_L^{(i)} - Yaw_R^{(i)}\right|$$

The standard deviation of difference for each angle is also determined by Eq. (5). And the results are shown in Table II.

$$\sigma_{Roll} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{Roll}^{(i)} - M_{Roll})^2}$$
$$\sigma_{Pitch} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{Pitch}^{(i)} - M_{Pitch})^2} \quad (5)$$
$$\sigma_{Yaw} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{Yaw}^{(i)} - M_{Yaw})^2}$$

TABLE II. THE MEAN AND STANDARD DEVIATION OF DIFFERENCE

Angles	Mean of difference (Degree)	Standard deviation of difference (Degree)
Roll	2.51^0	2.24^0
Pitch	4.84^0	3.35^0
Yaw	3.27^0	2.25^0
Average	$\mathbf{3.54^0}$	$\mathbf{2.61^0}$

The mean of difference of Roll, Pitch and Yaw angle is around only 2-5 degree as shown in Table II. And the average of three angles is around 3.54 degree while the standard deviation of difference is around 2.61 degree in average sense.

E. Processing Time (Latency)

The processing time (latency) of the proposed algorithm is determined from capturing raw images by the LMC to get the final hand pose after applying the coordinate transformation. It is divided into three parts including the processing time for image capturing, marker detection and coordinate transformation. The latency is determined and shown in Fig. 12. The latency of each part is also measured as shown in Table III using the timer function in Windows 10.

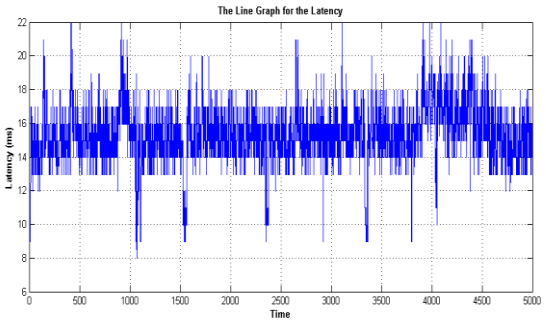


Figure 12. The line graph for the total latency

TABLE III. THE LATENCY RESULTS

Processing Step	Mean of latency (millisecond)
Image capturing	3.41
Marker detection	11.75
Coordinate transformation	0.075
Total	**14.96**

The total latency of the method is around 14.96 millisecond. The latency for marker detection is mainly, around 11.75 millisecond while the latency for coordinate transformation is only 0.075 millisecond in average sense. The experiments demonstrates successful tracking of user's hand pose in about 70Hz under VR-compatible PC environments.

V. CONCLUSIONS

A sensor fusion method is proposed to detect the pose of user's hand in HMD environments by using a LMC with simple circular markers on surfaces of a box wearing on the back of hand and two IMU sensors. One IMU sensor is located in the box and the other IMU sensor is fixed with the LMC. To use for VR/MR applications, MLP algorithm is adopted to transform the hand's pose in IMU coordinate system into the pose in LMC global coordinate system by minimizing MSE. The pose detection results are compared with poses of bare hand capturing by LMC. As the results,

the estimated data after transformation is fitted well with reference data by LMC in the sense that the average of mean difference for each Roll, Pitch and Yaw angle is around 3.54 degree. It is applied successfully to track and estimate the pose of user's hand in real-time, 70Hz under VR-compatible PC environments.

Future works include the improvement of the accuracy of hand pose estimation and the application for VR/MR in mobile smart phones.

REFERENCES

[1] HTC Corporation, "HTC VIVE Tracker Developer Guidelines v1.0", March 2018.

[2] J. L. Raheja, A. Chaudhary, K. Singal, "Tracking of Fingertips and Centres of Palm using KINECT", International Conference on Computational Intelligence, Modelling & Simulation, pp. 248-252, Sept. 2011.

[3] S. Ameur, A. B. Khalifa, M. S. Bouhlel, "A comprehensive leap motion database for hand gesture recognition", International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, pp. 514-519, Dec. 2016.

[4] C. Naidu1, A. Ghotkar, "Hand Gesture Recognition Using Leap Motion Controller". International Journal of Science and Research, vol. 5, Issue 10, October 2016

[5] M. Alimanova, S. Borambayeva, D. Kozhamzharo, "Gamification of Hand Rehabilitation Process Using Virtual Reality Tools: Using Leap Motion for Hand Rehabilitation", IEEE International Conference on Robotic Computing (IRC), pp. 336-339, April 2017.

[6] B. Khelil, H. Amiri, "Hand Gesture Recognition Using Leap Motion Controller for Recognition of Arabic Sign Language", 3rd International Conference on ACECS, PET, pp. 233-238, 2016.

[7] M. D. Wibowo, I. Nurtanio, A. A. Ilham, "Indonesian sign language recognition using leap motion controller", ICTS, pp. 67-72, Oct. 2017

[8] Q. Wang, Y. Wang, F. Liu, W. Zeng, "Hand Gesture Recognition of Arabic Numbers Using Leap Motion via Deterministic Learning", Chinese Control Conference (CCC), pp. 10823 – 10828, July 2017.

[9] CyberGlove Systems LLC "CyberGlove III MOCAP Glove User and Programmer Guide" published 2018.

[10] T. Baumann, T. Hao, Y. He, R. Shoda "Perception Neuron Unity Handbook", Perception Neuron Unity Integration 0.2.2, Noitom Technology Co., Ltd, June 2015.

[11] C. Y. Park ; J. H. Bae ; I. Moon "Development of wireless data glove for unrestricted upper-extremity rehabilitation system", 2009 ICCAS-SICE, pp. 790-793, Aug. 2009.

[12] T. L. Baldi, S. Scheggi, L. Meli, M. Mohammadi, D. Prattichizzo, "GESTO: A Glove for Enhanced Sensing and Touching Based on Inertial and Magnetic Sensors for Hand Tracking and Cutaneous Feedback", IEEE Transactions on Human-Machine Systems, pp. 1066-1076, Volume: 47 , Issue: 6 , Dec. 2017.

[13] B. Hartmann, N. Link, G. F. Trommer, "Indoor 3D Position Estimation Using Low-Cost Inertial Sensors and Marker-Based Video-Tracking", IEEE/ION Position, Location and Navigation Symposium, pp. 319-326, May 2010.

[14] T. K. Chan, Y. K. Yu, H. C. Kam, K. H. Wong, "Robust Hand Gesture Input Using Computer Vision, Inertial Measurement Unit (IMU) and Flex Sensors", IEEE International Conference on Mechatronics, Robotics and Automation (ICMRA), pp. 95-99, May 2018.

[15] Yong-Ho Lee, Mincheol Kim, Hwang-Youn Kim, Dongmyoung Lee, Bum-Jae You, "CHICAP: low-cost hand motion capture device using 3D magnetic sensors for manipulation of virtual objects", SIGGRAPH Emerging Technologies 2018, 4:1-4:2.

[16] Luc Cong Vu, Eun-Seok Choi, Bum-Jae You, "Marker Classification by Sensor Fusion for Hand Pose Tracking in HMD Environments using MLP", Korea Information Processing Society (KIPS) conference, South Korea, Nov. 2018, unpublished.

[17] G. Bishop and G. Welch, "An introduction to the Kalman filter" in SIGGRAPH Course Notes, Los Angeles, CA, 2001, pp. 1–81.