Wearable Visual-Inertial Hand Tracking Interface Regardless of Environment and Occlusion

Yongseok Lee, Hanbyeol Yoon, Jinwook Huh, Wonha Lee and Dongjun Lee

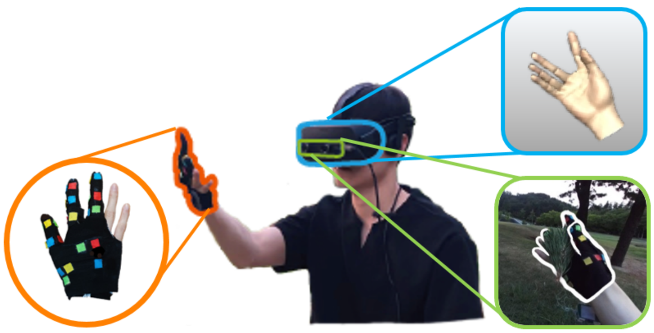

Fig. 1. System setup for our visual-inertial hand tracking system.

I. Introduction

As the higher mammals including human has the skeletal structure which consist of a large number of bones connected with joints and accordingly, many artificial system such as robot arm, humanoid also has imitated skeletal structure. Articulating pose of such systems is crucial but ongoing problem for many applications (e.g. control of humanoid, VR manipulation using natural user interface, whole body motion capture for SF movie, etc.)

Among these skeletal systems, human hand (e.g. [1], [2], [3], [4], [5]) and body (e.g. [6], [7], [8]) are mainly researched topics for their unlimited possibilities for many applications. One of the main ways to track the pose of human hand or body is based on vision-based techniques with the drastic advance in machine learning technology. However, the vision-based techniques have inherent limitations from the vision sensor, camera.

First of all, it fundamentally suffers from the occlusion problem since the ray from an object behind other object cannot reach camera lens. In case of tracking the skeletal structure with a large degrees of freedom, the self-occlusion where a link of system occludes another link frequently occurs, which vision-based tracking systems cannot capture. While machine-learning based method notably mitigates this occlusion problem by learning immense dataset including occlusion case (e.g. [5]), poses fundamentally impossible to be tracked by camera still exist, such as finger motion behind the dorsum of the hand.

Also, though this machine-learning based method shows the best performance ever among vision-based methods, this do not work well for not similar scene within training dataset. For example, environments which lack in existing datasets such as outdoor or factory, or hand wearing other devices such as gloves for sport, the disabled [9] or haptic device [1] will not work properly on the deep-learning based hand tracking. Moreover, in daily life, human hand ceaselessly interacts with a myriad of daily object with a variety of motion, however, any existing vision based method would not track hand motion with daily life object interaction due to the mentioned two problems : occlusion and dataset-only validity.

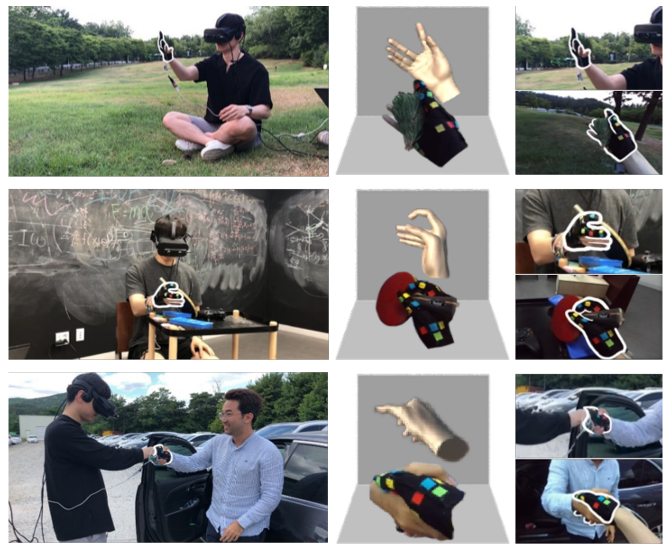

Fig. 2. Snapshots of qualtitative experiments : we verify robustness of our system under challenging scenarios, that is outdoor green field (top), object occlusion (middle) and outdoor human interaction (bottom).

To solve these issues of the vision-based system, there have been many wearable sensor based tracking systems, mainly utilizing IMUs or soft sensors, to track the large-DOF fingers motion. However, a few major issues of these hand tracking systems 1) are not stand alone, which means the global position of base hand should rely on external tracking system (e.g., HTC VIVE tracker, AR marker, Oculus sensor). 2) shows inferior accuracy from bias (e.g. accelerometer, stretch sensor), disturbance (e.g.linear acceleration, magnetic disturbance) or uncalibrated user-specific parameters (e.g. link length, sensor attachment offset). And 3) to estimate rotation of joint, magnetometer are widely used by providing global yaw direction. However, this sensor is vulnerable to change of environmental magnetic fields which needs an annoying calibration phase and sensitive to magnetic objects

This research was supported by a grant to Bio-Mimetic Robot Research Center Funded by Defense Acquisition Program Administration, and by Agency for Defense Development (UD190018ID).). Y. Lee, H. Yoon, J. Huh, W. Lee and D.J. Lee are with the Department of Mechanical and Aerospace Engineering of Seoul National University. Email: {yongseoklee, dbsgksquf13, jinukheo, whlee741, djlee}@snu.ac.kr

which surround our daily life.

Along these reasoning, we propose a novel visual-inertial tracking algorithm which utilizes both camera and IMU and optimally fuse complementary aspects of each sensor. By integrating merits of both sensors, our proposed method shows more accurate and robust performance under any circumstance. We develop our method as a general tracking frame work which is applicable to any skeletal structures. In this paper, we apply our method to hand tracking system in order to verify the performance and the applicability of the method. Consequently, our novel visual-inertial hand tracking system overcomes most of the issues which existing system suffer from. We verify from the following qualitative experiments in a few situations where existing hand tracking cannot work properly.

II. Overview of system

1) Hardware setup: Hardware of our hand tracking system is mainly composed of two parts: an wearable tracking glove and a stereo camera which provides visual information of point marker for sensor fusion.

Our new wearable tracking glove is developed based on our previous hardware in [1]. This glove has two types of sensors - MEMS IMU (Invensense MPU 9250®) and soft stretch sensor (StretchSense). Five IMUs are attached to five different IMU segments (i.e. dorsum of hand, metacarpal of thumb, and proximal phalanges of the three fingers (thumb, index, and middle finger)) and two stretch sensors are inserted inside the glove of the index and middle finger.

The way how we obtain the visual information of hand from the camera is based on point markers such as IR LED, pattern printings or deep-learning based feature. Specifically, in our hand tracking system, we adopt color blobs as point markers and attach them to the surface of a glove covering IMUs and stretch sensors. The color blobs are attached to the designated position of the glove including the position directly above the IMU. In detail, the color blobs are made of fabric with four different colors, and have square shape whose length of one side is 12mm. With thirty seven color blobs on the glove, the positional information of the attached marker can be obtained through triangulation of stereo camera, through the following algorithm in sec. II-.2.

As a vision sensor to observe the point marker set, a stereo camera is utilized. Specifically, our system utilizes ZED mini (Stereolabs) which is manufactured to be used with HMDs, namely, has proper weight, size and baseline for human usage (weight: 62.9g , Field of View: 90 (H) x 60 (V) x 110 (D), max. base line 63mm, resolution: 2560x720)

2) Algorithms: Our algorithm comprised of three steps : 1) point marker extraction, 2) point matching for correspondence search and 3) visual-inertial sensor fusion filter.

To explain each step briefly, in point marker extraction, we detect attached color blob based on HSV based thresholding. We set then minimal ROI from the current hand pose estimates to decrease computation cost or avoid extraction of blob not in our glove with the elimination of the pixel with too low saturation or value pixel. After simple noise elimination, we finally classify proper extracted contours in thresholding image as a blob, filtering by geometric condition such as size, solidity, circularity etc. Then the center of the contour is determined as a 2D position of blob Measurement.

In the point matching stage, we perform two sequential matching process that is, 1) 2D stereo marker matching which recovers 3D position of marker from 2D marker positions in both right and left camera image and 2) 3D correspondence search of 3D marker position measurement with true marker on FTM. Through this process, each marker on the glove are matched with the recovered 3D positions of point marker from stereo camera.

In the sensor fusion filter, we finally estimate the hand pose, sensor bias, and user parameter. Our system model needs real-time optimization for a large number of states, so we adopt EKF as an estimator for vision and wearable sensor fusion. Mainly the IMU information, i.e. information of accelerometer and gyroscope, are utilized to predict next-step hand pose and then this propagated hand pose are corrected by visual information of point marker set from the point matching stage. Rich and robust visual information from the point matching can calibrate wearable sensors bias and user-specic parameter such as IMU attachment offset or hand scale, which all enhance both performance and usability of the hand tracking system.

III. Experiments

Our system can track a hand robustly in the environment where existing hand tracking system cannot operate stably. The qualitative experiments are performed to show the robustness of our system, the outstanding results of which is shown in the fig. I. Our system successfully perform tracking under the four challenging cases: 1) hands on the colorful background, 2) occluded hand configuration, 3) hands on the outdoor environment, 4) hands equipped with haptic device, which will also be presented by video and demo.

IV. Live Demonstration Plan

We plan to show the performance of our tracking system by letting users experience different kinds of scenarios. Firstly, the visitors can utilize our hand tracking system while manipulating a virtual object and playing custom-built game in VR using their hands. Such experience can make them feel vivid interaction through VR, and realize the broad prosperity of hand tracking systems in the field of VR. On top of that, the users can get an opportunity to interact with daily objects when our system track their hands correctly and robustly. The robustness of our tracking system to occlusion is emphasized with this scenario.

For that, We prepare the hardware of our visual-inertial hand tracking system - the wearable tracking glove, the stereo camera - and other supplementary device including Occulus rift HMD and laptop. To be specific, we make three tracking gloves in different sizes in order for our system to be available for a wide range of hand sizes. By using simple and mobile hardware setup, we also can show the mobility of our tracking system.

REFERENCES

[1] Yongjun Lee, Myungsin Kim, Yongseok Lee, Junghan Kwon, Yong-Lae Park, and Dongjun Lee. Wearable finger tracking and cutaneous haptic interface with soft sensors for multi-fingered virtual manipulation. *IEEE/ASME Transactions on Mechatronics*, 24(1):67–77, 2019.

[2] Tommaso Lisini Baldi, Stefano Scheggi, Leonardo Meli, Mostafa Mohammadi, and Domenico 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*, 47(6):1066–1076, 2017.

[3] Chen Qian, Xiao Sun, Yichen Wei, Xiaoou Tang, and Jian Sun. Realtime and robust hand tracking from depth. In *Proc. of IEEE conference on computer vision and pattern recognition*, pages 1106–1113, 2014.

[4] Gyeongsik Moon, Ju Yong Chang, and Kyoung Mu Lee. V2v-posenet: Voxel-to-voxel prediction network for accurate 3d hand and human pose estimation from a single depth map. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 5079–5088, 2018.

[5] Franziska Mueller, Florian Bernard, Oleksandr Sotnychenko, Dushyant Mehta, Srinath Sridhar, Dan Casas, and Christian Theobalt. Ganerated hands for real-time 3d hand tracking from monocular rgb. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 49–59, 2018.

[6] Jan Beutel, Bernhard Buchli, Federico Ferrari, Matthias Keller, Marco Zimmerling, and Lothar Thiele. X-sense: Sensing in extreme environments. In *Proc. of the Design, Automation & Test in Europe*, pages 1–6, 2011.

[7] Thomas Probst, Andrea Fossati, and Luc Van Gool. Combining human body shape and pose estimation for robust upper body tracking using a depth sensor. In *Proc. of the European Conference on Computer Vision*, pages 285–301, 2016.

[8] Francesc Moreno-Noguer. 3d human pose estimation from a single image via distance matrix regression. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2823–2832, 2017.

[9] Hyunki In, Brian Byunghyun Kang, Minki Sin, and Kyu-Jin Cho. Exo-glove: A wearable robot for the hand with a soft tendon routing system. *IEEE Robotics & Automation Magazine*, 22(1):97–105, 2015.