AN EFFICIENT FUSION SCHEME FOR HUMAN HAND TRAJECTORY RECONSTRUCTION USING INERTIAL MEASUREMENT UNIT AND KINECT CAMERA

An Undergraduate Research Scholars Thesis

by

TRUNG LE

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisor: Dr. Roozbeh Jafari

May 2017

Major: Electrical Engineering

TABLE OF CONTENTS

ABSTRACT

An Efficient Sensor Fusion Scheme for Human Hand Trajectory Reconstruction Using Inertial Measurement Unit and Kinect Camera

Trung Le
Department of Electrical and Computer Engineering
Texas A&M University

Research Advisor: Dr. Roozbeh Jafari
Department of Biomedical Engineering
Texas A&M University

The turn of 21st century has witnessed an evolving trend in wearable devices research and improvements in human-computer interfaces. In such systems, position information of human hands in 3-D space has become extremely important as various applications require knowledge of user's hand position. A promising example of which is a wearable ring that can naturally and ubiquitously reconstruct handwriting based on motion of human hand in an indoor environment. A common approach is to exploit the portability and affordability of commercially available inertial measurement units (IMU). However, these IMUs suffer from drift errors accumulated by double integration of acceleration readings. This process accrues intrinsic errors coming from sensor's sensitivity, factory bias, thermal noise, etc., which result in large deviation from position's ground truth over time. Other approaches utilize optical sensors for better position estimation, but these sensors suffer from occlusion and environment lighting conditions. In this thesis, we first present techniques to calibrate IMU, minimizing undesired effects of intrinsic imperfection resided within cheap MEMS sensors. We then introduce a Kalman filter-based fusion scheme incorporating data collected from IMU and Kinect camera, which is shown to overcome each sensor's disadvantages and improve the overall quality of reconstructed trajectory of human hands.

DEDICATION

To my family, who have always been on my side throughout this journey.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my deepest gratitude to my research advisor, Dr. Roozbeh Jafari, who has provided me invaluable advice along the course of this research, and extended all his resources to support my endeavors.

I also would like to convey my sincere appreciation for Ph.D. students in ESP lab: Jian Wu, Ali Akbari, and Xien Thomas for continuously providing their insights and working closely with me to resolve challenges throughout the research.

NOMENCLATURE

IMU	Inertial Measurement Unit
MEMS	Micro Electro-Mechanical System
HCI	Human-Computer Interface
DOF	Degree of Freedom

CHAPTER I

INTRODUCTION

As computers become more powerful in computational capability and more portable in size, they have been introduced in almost every aspect of modern society. Interacting with computers in form of smart devices has become an essential part of daily life. However, current human-computer interaction (HCI) models have not caught up with the rapid improvements in capabilities of such devices. Old interaction modalities such as mouse and keyboards are becoming obsolete, inappropriate and troublesome for the changes in hardware design of smart devices, which are now becoming wearable and universal. Among the modern HCI systems, gesture based platforms prevail especially in wearable devices, where interaction with such devices needs to be natural, non-invasive and convenient. These platforms also proved their supremacy as they are fused with the wearable system themselves, and do not require many computational resources. In the wake of such systems, there is an increasing need for techniques and models to keep track of the motion of human hand during writing activity, and reconstruct handwriting as a useful input to such systems, since text is among the most intuitive and common method for human to interact with computer, or with each other with computer as an intermediary. Writing is a natural form for inputting texts which a majority of population can perform and is considered a human instinct. In wearable systems where devices need to be portable in terms of physical size, and interaction with the devices need to be subtle so as to not distract other people, one can’t use touchscreens or voice commands – the two popular interfaces in modern electronic devices – to interact with such wearable systems. Quick and short handwriting commands can be thought of as an alternative in those instances to control the device and compose texts. More interestingly, these handwritings can even be created without an external stylus. The user can simply wear a ring and write with his finger while the ring collects information of the finger movements during writing phase. There are also situations in which original handwriting could be of interest to be recorded, digitized and stored for later use. For an example, a professor wants to use traditional chalks and blackboard for his lectures as it is an effective way of teaching, but still wishes what he writes on the board to be recorded and electronically stored for later references, or distributed to his students as class notes. A wearable ring which keeps track of the professor’s hand motion and reconstructs his handwriting from the recorded data is also a potential solution for this case.

Unfortunately, attempts by researchers to build such system encountered major hurdles that prevent it from functioning reliably. Apart from technical difficulties driven by customer’s needs, e.g. low power consumption, reliable signal transmission, portable and aesthetic form, etc., there are also adversities from within the nature of modalities used. A commonly used motion sensor is IMU because of its compact form and low cost, which is extremely important to encourage users to wear comfortably over a long period of time. An IMU consists of 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer. There is no direct way to determine position from the above sensors. In order to derive position, one must integrate the accelerometer readings over time to get velocity, and then integrate velocity one more time to get position. This method undergoes dead reckoning problem, where current position is obtained from a previously determined position, and is used to get the next position with the knowledge of speed. The sequence of position that is based and advanced upon each other can easily deviate from ground truth if a small error is introduced at one certain step, as this error would linger and have an effect on all subsequent steps. Through time this error will become substantially large and the results can no longer be trusted if there is no additional reference point introduced to correct this calculation.

To avoid the dead reckoning problem, other modalities which can measure directly absolute position are used. Optical sensors, i.e. cameras, are widely used for object tracking in fixed and confined environment (rooms, buildings). Depth camera, which adds infrared projector to measure space depth, has extended the position estimation capability of traditional cameras from 2-D to 3-D space. However, cameras do not perform well when tracked object is hidden behind other objects, or environment lighting condition is poor.

In this thesis we solve the problem of tracking hand position by proposing a hybrid system incorporating both IMU and Kinect depth camera. The structure of the thesis will be as follows: first we will introduce the design of hardware system used; next we will discuss signal processing techniques implemented to minimize the intrinsic errors of the two sensors, following by a Kalman filter-based fusion scheme taking into account the orientation differences of each sensor. We then present the experimental design, discuss the results and suggest future works.

CHAPTER II

RELATED WORKS

Human gestures and body movements have recently gained notable attention as a prominent source for modern human-machine interface thanks to its wide applications in rehabilitation, virtual reality, robotics, health monitoring, etc. These applications often come in form of wearable devices and take huge advantage of low-cost, microelectromechanical system (MEMS) inertial measurement units (IMU) [2 - 4]. A typical IMU consists of accelerometer, gyroscope and magnetometer, which can be used selectively to derive motion and orientation of human body parts where they sensor is worn. In [7], Gummeson et al presented an efficient ring for capturing human gestures on surface. By using acoustic sensing integrated with inertial sensing, the ring was able to capture finger motion when writing on surface and recognized a set of 12 stroke-based gestures for text entry. Handwriting recognition as targeted in this paper or hand gestures recognition as in many other ones [5, 6], however, is not as vulnerable to sensors' errors as in handwriting reconstruction problem. Pen-based input instruments also drew attention from many researchers [11 - 13]. Wang et al in [11] introduced an IMU-based pen together with their proposed MAD switch algorithm in an effort to reliably reconstruct handwriting from data collected by accelerometer and gyroscope. Their work used the Zero Velocity Compensation technique first introduced in [13] by Bang et al, which minimize drift effect by resetting velocity to zero after each motion. However, attempts to reconstruct handwriting (or motion trajectory in other words) solely from IMU would still suffer from drift error although there were corrective techniques involved to limit the errors within individual strokes (the beginning and end of strokes are determined by observing if acceleration and angular velocity exceed/go below certain threshold). Zhang et al in [12] presented an IMU pen with a Kalman filtering algorithm and showed that such filter can minimize the inherent noise of IMU sensor. Tsang et al in [14] incorporated an electromagnetic resonance (EMR) motion detection board as an additional modality to get direct position of writing strokes as their IMU pen slides on the board's surface. Kalman filter was used in their work to fuse the two modalities (IMU pen and EMR board) and showed an improvement in the quality of reconstructed trajectory.

Others have shifted approaches to vision-based system. As depth cameras like Kinect become commercialized and provide accurate position estimation of human body joints, they have been utilized extensively in recent human body tracking systems [15 – 17]. In [16], Destelle et al promoted Kinect for hand detection and hand gesture recognition. Gabel et al in [17] utilized Kinect to develop a system for full gait analysis which could extract stride intervals and arms kinematics. In [9] Frati and Prattichizzo combined Kinect with wearable haptic device for hand tracking application. It is known for such vision-based approach that the tracking result obtained from solely Kinect camera would suffer from occlusion, i.e. when the interested human body part is hidden by other objects.

It is natural thinking to combine the two approaches and come up with a sensor fusion paradigm to take advantages of both IMU and Kinect as presented in [10]. Such novel direction requires an implementation of Kalman filter as the underlying theoretical framework for data fusion from multiple modalities. In the following sections, we will discuss in detail the design of our wearable system utilizing IMU and Kinect, backed by Kalman filtering algorithm and aimed to apply specifically to handwriting reconstruction.

CHAPTER III

SYSTEM DESIGN

As mentioned in previous sections, IMU and Kinect are two fundamental components of our system. The two elements will be integrated under Kalman filtering algorithm to harness each sensor's advantages and improve the reconstructed trajectory of human hand wearing the IMU (Fig.1).

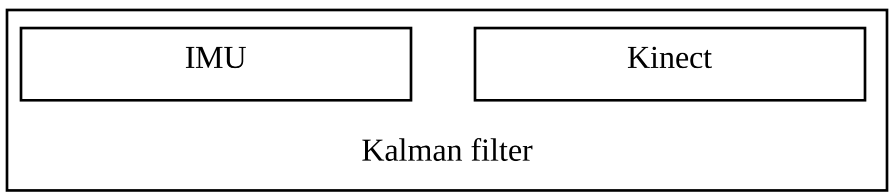

Fig. 1: Overall system design with Kalman filter as underlying framework

IMU is a motion sensor commonly used across applications including navigation systems, gaming devices and wearable platforms. A typical MEMS (micro electromechanical system) IMU consists of 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer. Accelerometer measures acceleration of the IMU. This acceleration is the vector sum of motion acceleration and gravitational acceleration. Gyroscope measures angular velocity around each axis. Magnetometer measures the Earth magnetic field on each axis. In this thesis, the magnetometer was not used.

MotionNet platform

The IMU is integrated on MotionNet board, a motion sensor platform developed in our lab (Fig. 2 and 3). The board consists of 9-DOF IMU, analog input interface, dual mode Bluetooth module, power management and charging circuit, microSD card module. The center microcontroller is low-powered MSP430 manufactured by Texas Instruments. Data collected by accelerometer, gyroscope and magnetometer is sent via Bluetooth to host computer, on which subsequent signal processing steps are performed. The platform is confined in a plastic box which is worn at the user's wrist during experiment.

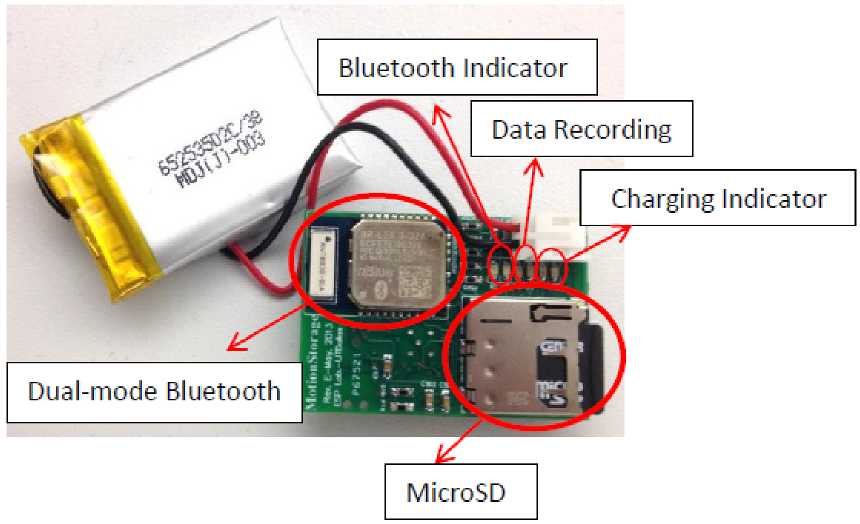

Fig. 2: Front view of MotionNet platform [6]

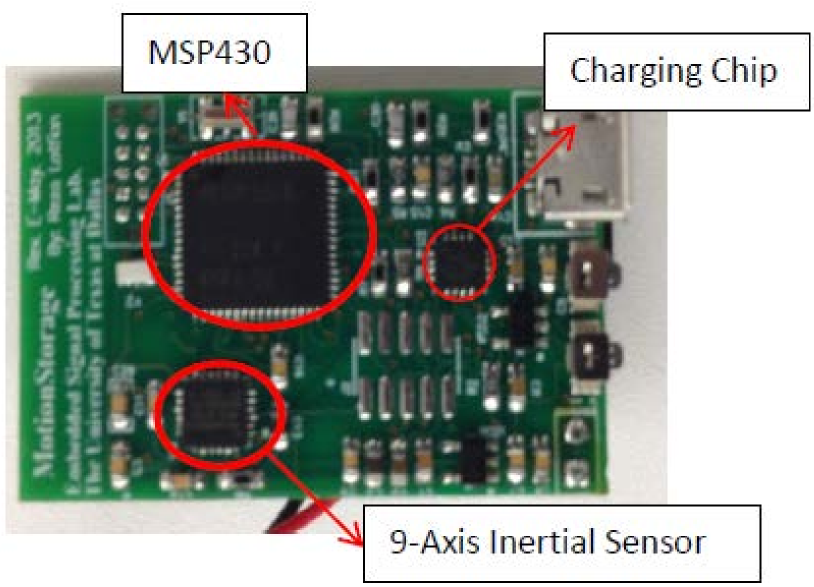

Fig. 3: Bottom view of MotionNet platform [6]

Kinect

Kinect (Fig. 4) is a motion sensing input device initially developed by Microsoft to support the company's Xbox 360 video game consoles. It is later widely used by researchers across various applications. The hardware includes an RGB camera and a depth sensor. The depth sensor consists of an infrared laser projector and monochrome CMOS sensor, which creates a grid to the front of the camera so that position of object within the range could be determined precisely. Kinect is shipped with a Software Development Kit (SDK) for developers to interact with Kinect hardware. In our lab, we used this SDK to extract human joints position. The right wrist joint was selected thanks to its reported better accuracy compared to left wrist [1]. Therefore the MotionNet platform was worn around the right wrist and the handwriting task was performed by assuring right hand (holding pen) and right forearm always in a stationary position with respect to each other.

Fig. 4: Microsoft Kinect 2 [18]

In terms of signal processing tools, MATLAB was chosen for the program's myriad of efficient built-in functions and ease of working with large data sets thanks to its emphasis on matrix data representation.

CHAPTER IV

TRAJECTORY RECONSTRUCTION

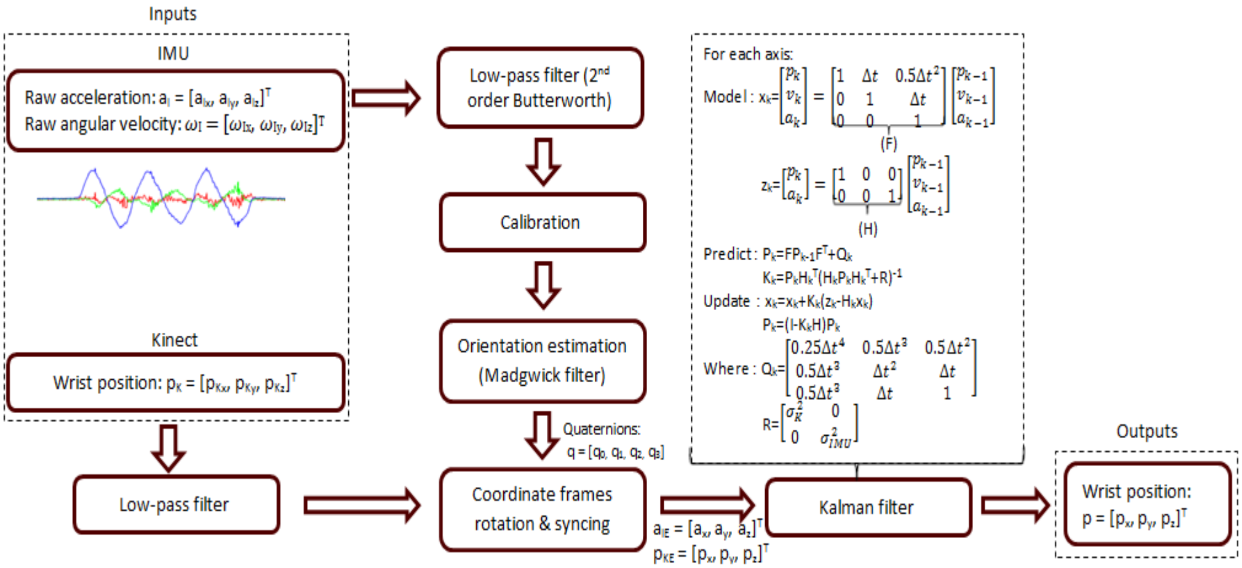

Fig. 5: Overall trajectory reconstruction flow diagram

Our overall design can be broken down into six stages as shown in Fig. 5: Low-pass filtering, calibration, orientation estimation, coordinate frames rotation and gravity removal, data syncing and Kalman filtering. Details of each stage are discussed in the sections below.

Low-Pass filtering

Raw readings from accelerometer, gyroscope and Kinect are affected by various unknown factors such as thermal noise, which would distort the real signal. In order to filter out these noise, first we assume this noise to reside in high frequency band and therefore a 2nd order Butterworth low-pass filter is utilized to smoothen the readings. The practically best cutoff frequencies for acceleration, gyroscope and Kinect were found based on experimentation to be 6 Hz, 6 Hz, 2 Hz, respectively.

Calibration of accelerometer and gyroscope

Commercial MEMS IMUs are cheap and therefore are usually not fine-tuned to a high level of accuracy. Due to manufacturing processes, impacts of working temperature and inherent degradation over time, raw data obtained from IMU is subjected to certain offset and scale factor. If not properly calibrated, IMU would yield incorrect readings and thus introduce a large amount of error to our trajectory reconstruction algorithm. For accelerometer calibration, in this thesis we use the calibration technique developed by Frosio et al as presented in [8].

Acceleration readings subjected to factory offset and scale factor can be represented by:

$$A = S(V - O) \qquad (1)$$

Where V is the true readings, $A = \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix}$ is acceleration read from the uncalibrated accelerometer, $O = \begin{bmatrix} O_x \\ O_y \\ O_z \end{bmatrix}$ is factory offset on each axis, $S = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix}$ where diagonal is scale factor on each axis and the other elements are cross-axis scale factors.

The assumption for correct acceleration readings is that the vector sum of acceleration on three axes must be equal to gravity g if the sensor is static. The error due to imperfect sensor is therefore:

$$e = a_x^2 + a_x^2 + a_x^2 - g^2 = \sum_{i=x,y,z}\{\sum_{j=x,y,z}\left[S_{ij}\left(V_j - O_j\right)\right]^2\} - g^2 \quad [8] \qquad (2)$$

Cummulative error measured over all (at least 9) orientations is:

$$E = E\left(O_x, O_y, O_z, S_{xx}, S_{yy}, S_{zz}, S_{xy}, S_{xz}, S_{yz}\right) = \frac{\sum_{k=1}^{N} e_k^2}{N} \quad [8] \qquad (3)$$

This error can be minimized with respect to parameters S and O using Newton’s method, a nonlinear optimization algorithm. This method is applied iteratively until convergence, at which E is minimized and values of S and O are obtained.

For gyroscope, the calibration procedure is as follows. Assuming the scale factor of gyroscope is negligible; offset of gyroscope on each axis is the only major contribution to inherent angular velocity error. We then calculate the mean of first 200 samples angular velocity in static position. The reading obtained should be zero. However due to factory offset it is different from zero, therefore we consider this mean to be the factory offset and subtract this value for all gyroscope readings to get the correct readings.

Orientation estimation

It is our great interest to know the orientation of IMU during the motion. There are two reasons for this desire:

First, accelerometer always measure gravity acceleration. Although this is important for static orientation estimation, we only care about acceleration caused by the hand motion itself. Gravity acceleration always sums up with motion acceleration, thus the obtained final readings do not reflect exactly the motion acceleration, which is the only thing we want. Knowing the orientation of IMU can help us rotate the IMU to the sensor Earth frame (where g always equal to 1 on the z axis) and subtract the gravity from there to get pure motion acceleration.

Second, hand motion is performed freely in 3-D space. As human hand moves, the orientation of IMU also changes accordingly. Even if the user tries to fix the orientation of IMU during motion, subtle vibration from his/her hand could also unintentionally change the orientation of IMU. We want to know the orientation of IMU at every instance of time so as to have proper adjustment to acceleration readings (affected by orientation) in each time step.

In this thesis, we utilize the efficient orientation estimation method developed by Madgwick [19]. Madgwick filter is a computationally efficient gradient descent-based optimization method used to obtain quaternion representation of IMU. The filter version for 6-

DOF IMU is used in this thesis, where only acceleration and angular velocity are used as inputs to get orientation of sensor's body frame with respect to the sensor's Earth frame. The Madgwick filter outputs estimated orientation in the form of quaternions. Quaternion is a four-dimensional complex number$q = [q_0, q_1, q_2, q_3]$, which was introduced as a solution to the Gimbal Lock problem as it allows an additional dimension to represent orientation of a rigid body.

Quaternion representation also comes in handy when it is our turn to sync different sensor's coordinate systems later on. We can get an orientation of frame B relative to frame A (denoted by ${}^A_Bq$) via a rotation of angle θ around arbitrary axis r in frame A:

$$ {}^A_Bq = [q_1 \quad q_2 \quad q_3 \quad q_4] = [\cos\frac{\theta}{2} \quad -r_x \sin\frac{\theta}{2} \quad -r_y \sin\frac{\theta}{2} \quad -r_z \sin\frac{\theta}{2}] \qquad [18] \qquad (4) $$

Coordinate frames rotation and gravity removal

Kinect and IMU each has each own body frame as shown in Fig. 6:

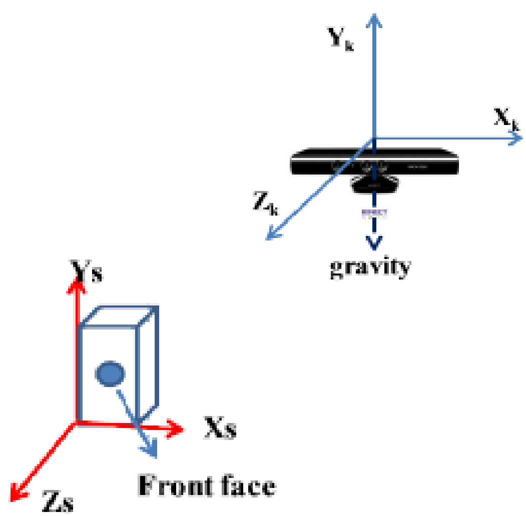

Fig. 6: Body coordinate systems of Kinect and IMU [20]

As discussed in the previous sections, IMU coordinate systems is freely rotated in space as the user moves his/her hand. This makes it hard for us to incorporate the two sensors readings as they are simply not "talking in the same language". In order for the fusion scheme to work, data recorded by both sensors must represent the motion in a common coordinate system. Since Kinect is always affixed in space, we need to rotate the readings of IMU to align with the Kinect's coordinate frame (which is also considered the global frame in our work). The result is the IMU readings represented in IMU's Earth frame. In the scope of this research, it is still of user's responsibility to keep axes of the IMU Earth frame parallel to axes of Kinect frame during the motion, although the axes can switch places with each other. Later on we only feed acceleration data to the Kalman filter, thus we only need to rotate the acceleration obtained by IMU to the IMU's Earth frame. The rotation can be done by multiplying the acceleration with the rotation matrix T:

$$A_E = TA_B \qquad (5)$$

Where A_E and A_B are accelerations in IMU's Earth frame and body frame, respectively, and $T = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 + q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}$

Reminder: A is 3x1 matrix, $A = \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix}$, and $q = [q_0, q_1, q_2, q_3]$ is obtained from the output of Madgwick filter.

After rotating IMU to its Earth frame, we have isolated gravity component to exist only in z direction. From now the pure motion acceleration can be obtained by subtracting gravity in z axis:

$$A_{motion} = A_{rotated} - g = \begin{bmatrix} A_{x,rotated} \\ A_{y,rotated} \\ A_{z,rotated} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad (6)$$

Next, Earth frame of IMU (where z is pointed downward and x, y of IMU parallel to x, z of Kinect) needs to be rotated to match Kinect frame (also considered global frame in our work) by the following axis transformations:

$$A_{\text{motion, x global}} = - A_{\text{motion, x local}} \quad (7)$$

$$A_{\text{motion, y global}} = - A_{\text{motion, z local}} \quad (8)$$

$$A_{\text{motion, z global}} = A_{\text{motion, y local}} \quad (9)$$

As of now we successfully have the two reference frames match each other.

Data synchronization

IMU has a sampling rate of 200 MHz, while Kinect has sampling rate of 30 Hz, much lower than IMU. It is our interest to keep the sampling rate of IMU, since it is necessary to not miss rapid movements due to inadequate sampling rate. But at the same time we need the two sensors to have a same sampling rate for data processing. We “upsampling” Kinect by replicate preceeded sample point in time steps where IMU data is present but Kinect data is missing. As a result these replicated samples are what we inferred rather than measured directly by Kinect, thus there should be a certain uncertainty associated with these samples. This uncertainty would be best represented by assigning a high value of covariance in measurement noise covariance matrix R (see Kalman filtering section) for Kinect samples in these time periods.

The method of assigning a high level of uncertainty for Kinect whenever confident measurement from Kinect is missing will also help in cases of occlusion. If occlusion happens, Kinect won’t be sending data of wrist’s position. As data from Kinect is missing, according to our data syncing scheme described above, wrist’s position in this period would take the value of last tracked position, but is associated with a high measurement noise in R. Therefore, we would see in the Results section that our fusion’s position would rely more on position estimated by

IMU in these time intervals. In this way, IMU proved to be useful whenever we have outage of Kinect, i.e. occlusion happens.

Kalman filtering

Kalman filter is an optimized filter, which takes the information of noise and other inaccuracies into account with a series of observable measurements over time, and outputs more accurate estimated values of unknown variables. Kalman filter is widely used especially in sensor fusion problems.

The overall Kalman filtering algorithm for our specific application is presented as follow:

For each axis:

Model : $$x_k=\begin{bmatrix} p_k \\ v_k \\ a_k \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & 0.5\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{k-1} \\ v_{k-1} \\ a_{k-1} \end{bmatrix} \quad (10)$$

(F)

$$z_k=\begin{bmatrix} p_k \\ a_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{k-1} \\ v_{k-1} \\ a_{k-1} \end{bmatrix} \quad (11)$$

(H)

Predict : $P_k=FP_{k-1}F^T+Q_k$ (12)

$K_k=P_kH_k^T(H_kP_kH_k^T+R)^{-1}$ (13)

Update : $x_k=x_k+K_k(z_k-H_kx_k)$ (14)

$P_k=(I-K_kH)P_k$ (15)

State vector:

Based on our application of trajectory reconstruction, we define our state x to have information of position, velocity and acceleration, $x = [p, v, a]^T$

State transition equations:

$a_k = a_{k-1}$ (16)

$v_k = v_{k-1} + \Delta t \; a_{k-1}$ (17)

$p_k = p_{k-1} + \Delta t \; v_{k-1} + 0.5\Delta t^2 a_{k-1}$ (18)

We will model acceleration as having associated process noise, which makes a_k not equal to a_{k-1}.

Thanks to linear algebra, transition from previous state to current state can be realized with the help of state transition matrix F = $\begin{bmatrix} 1 & \Delta t & 0.5\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}$

Measurement vector:

There are two measurements being made: position of subject's wrist measured by Kinect and acceleration of wrist measured by IMU. Our measurement vector is thus: $z = [p, a]^T$. The mapping between current measurement and previous state is expressed via matrix H=$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Process noise covariance matrix:

Process noise covariance matrix depicts how process noise would affect the prediction error, Q_k=$\begin{bmatrix} 0.25\Delta t^4 & 0.5\Delta t^3 & 0.5\Delta t^2 \\ 0.5\Delta t^3 & \Delta t^2 & \Delta t \\ 0.5\Delta t^3 & \Delta t & 1 \end{bmatrix} \sigma^2_{acc,IMU}$

Measurement noise covariance matrix:

Measurement noise covariance matrix depicts how measurement noise would affect the gain of Kalman filter, R=$\begin{bmatrix} \sigma^2_{Kinect} & 0 \\ 0 & \sigma^2_{acc,IMU} \end{bmatrix}$. This matrix would help determine which sensor is more reliable for Kalman result to follow at each instance of time.

Standard deviation of Kinect and IMU is measured on the first 200 samples. As discussed in Data Synchronization section, standard deviation of Kinect is set to infinity at time steps where Kinect does not have reliable measurements due to the lower sampling rate compared to IMU, or whenver occlusion takes place.

Initialization:

The state vector is initialized as $x = [p_{0,Kinect}, 0,0]^T$, where $p_{0,Kinect}$ is the initial position of wrist measured by Kinect. Error matrix P is initialized as 3x3 identity matrix.

Next state prediction:

In this stage, current state vector is predicted from previous state based on state transition equations. Error matrix P is computed based on the process noise and previous error matrix, as in equations (12). This new P is then used to calculate Kalman gain K as in equation (13), taking current measurement noise int account.

Next state update:

Newly computed Kalman gain will be used to decide whether the filter woud favor current measurement or the previous state more in order to update its state vector (equation (14)). If current measurement noise is large, i.e. measurement is not reliable as told by Kalman gain, then the filter will rely on previous state to estimate the current state and vice versa. New error matrix is also updated based on the Kalman gain (equation (15)).

CHAPTER V

RESULTS

The experiments were carried out as follows: A human subject worn MotionNet platform on his right wrist and sat in front of Kinect camera. His wrist and IMU are innitially hold static for 30 seconds to ensure adequate time for Madgwick filter to converge. He then moved his hand on a horizontal surface (xz plane) as if he was writing something. In order to assess the accuracy of the fusion scheme, he would trace his hands along fixed paths of various shapes. We investigated the effectiveness of the fusion scheme under two circumstances: when there is no occlusion and when occlusion happens.

No occlusion

The designated paths presented below (Fig. 7 a-e) were a rectangle of US letter size (21.6 x 28 cm). Data was recorded from Kinect and IMU and then fed to fusion algorithm to reconstruct his hand trajectory. The results for this shape are as shown below:

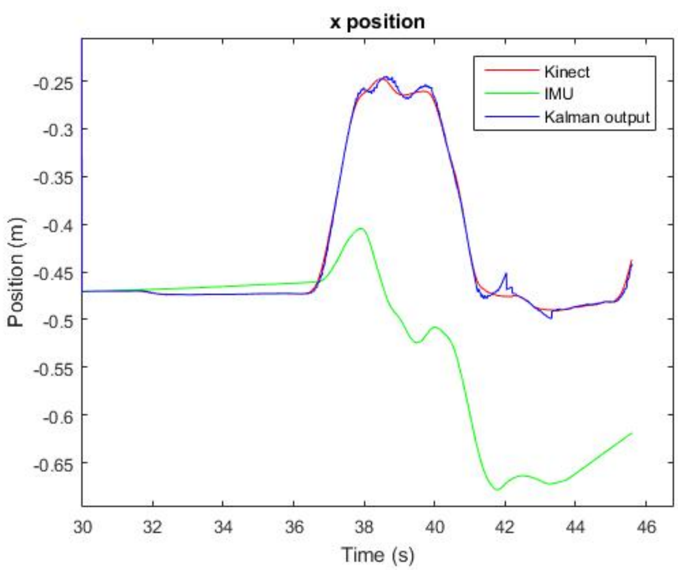

Fig. 7a: x position estimated by IMU, Kinect and Kalman filter for rectangular shape

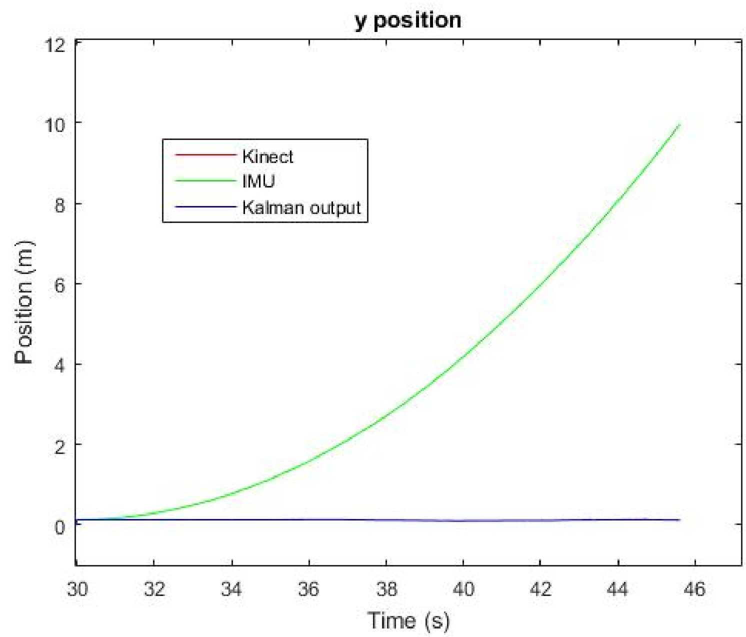

Fig. 7b: y position estimated by IMU, Kinect and Kalman filter for rectangular shape

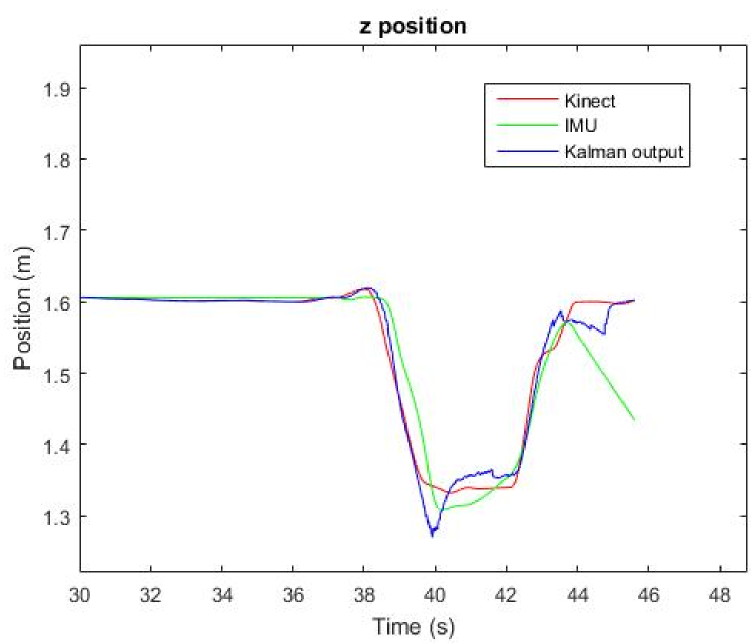

Fig. 7c: z position estimated by IMU, Kinect and Kalman filter for rectangular shape

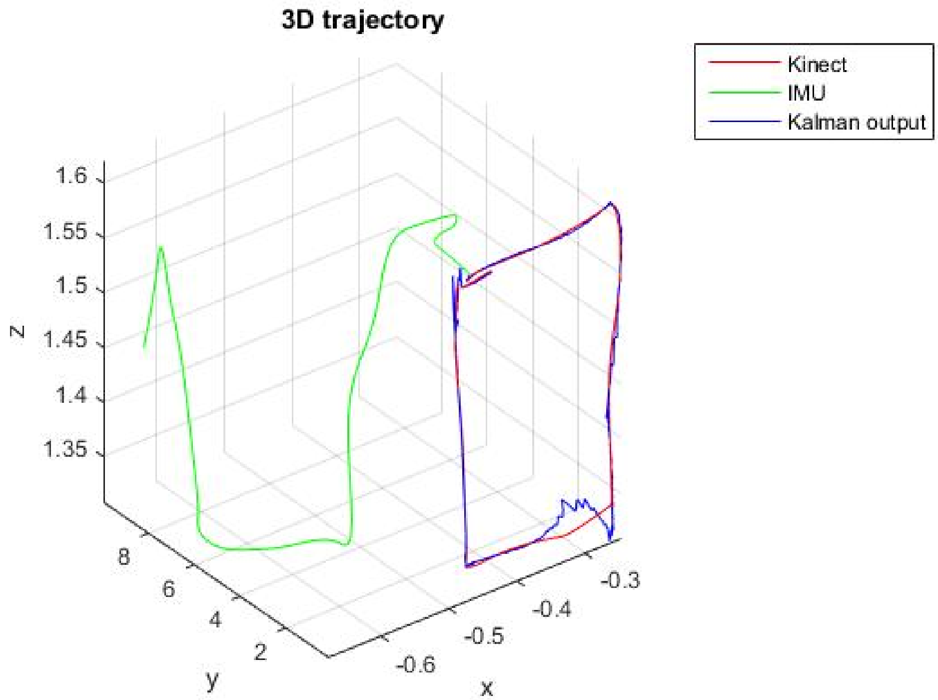

Fig. 7d: 3-D reconstructed trajectory by IMU, Kinect and Kalman filter for rectangular shape

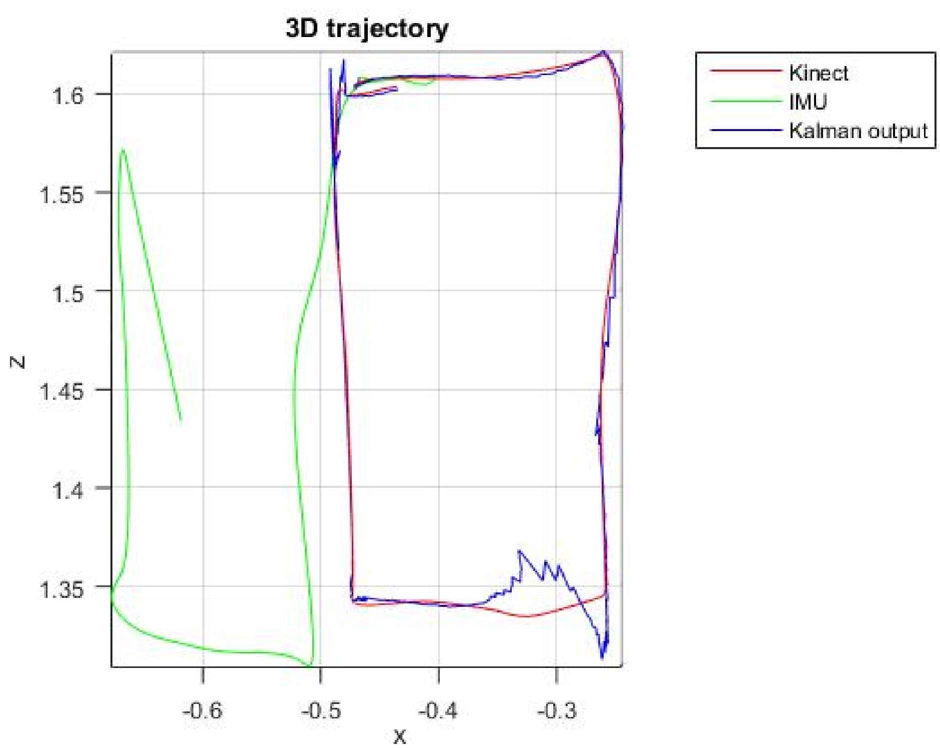

Fig. 7e: Reconstructed trajectory by IMU, Kinect and Kalman filter projected onto xz plane

In Fig. 7 a-e, the red lines represent position estimated by Kinect, green lines represent position estimation done by double integration of IMU's acceleration, and blue lines represent position estimated by the fusion scheme. It can easily be seen that trajectory reconstructed solely by IMU suffers significantly from drift error, as this is obtained via double integrating acceleration readings, while the estimation done by Kinect better resembles true rectangle. Thus the fusion scheme should prefer Kinect over IMU whenever Kinect signal is available. We can indeed deliberately set covariance of Kinect to be much less than that of IMU so that Kalman filter would follow Kinect estimation rather than IMU (Fig. 7d, e).

For the shape of rectangle, the reconstructed trajectory is shown to be bound by a rectangular of roughly 24 x 27 cm (error of ~2 cm on each axis compared to US letter size of 21.6 x 28 cm). We leave it for future works to set up a ground truth system (e.g. Vicon motion capture system [5]) in order to evaluate quantitatively the accuracy of each technique.

Experiments with other shapes were also carried out (Fig. 8 and 9), all confined by a box of US letter size.

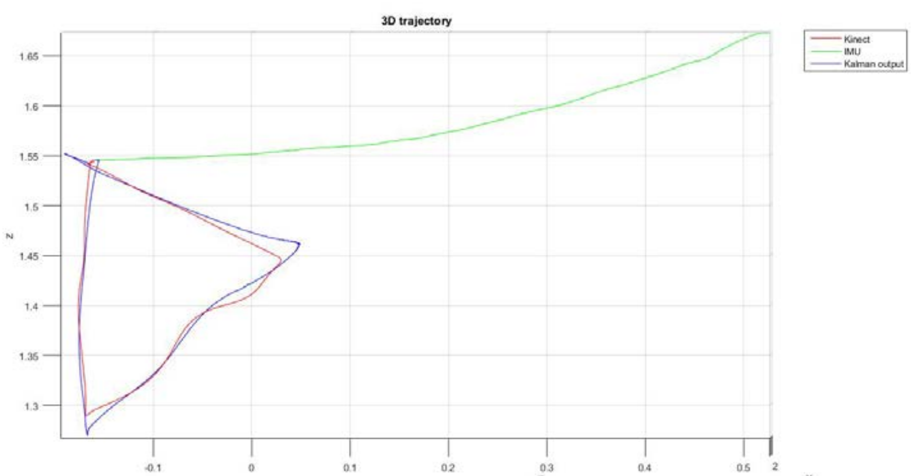

Fig. 8: 3-D reconstructed trajectory of a triangle confined in a US letter sheet

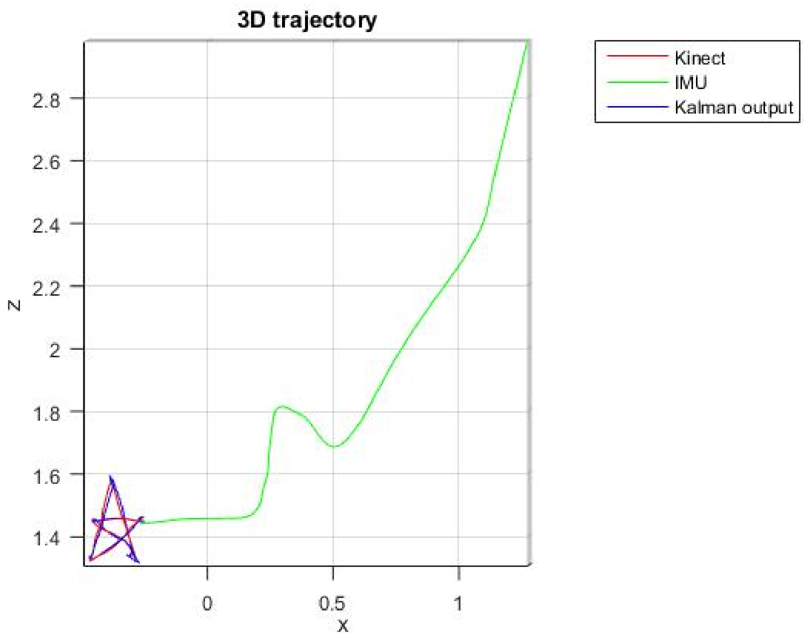

Fig. 9: 3-D reconstructed trajectory of a star confined in a US letter sheet

Occlusion

As discussed in Chapter IV, Kinect fares better than IMU in terms of position estimation because unlike IMU, Kinect does not accumulate errors through double integration of noisy acceleration readings to obtain position. Instead, position can be directly measured by depth camera. However, Kinect has lower sampling rate and suffers from occlusion, i.e. when line of sight from Kinect to human body is interrupted by other objects. Under these circumstances, Kinect will not be able to track subject's hand and thus will not send any data ("Kinect outage"). In our design, in order to synchronize Kinect's data length with that of IMU, Kinect is made to hold value of the last tracked position associated with a high uncertainty in periods of time when Kinect data is missing due to occlusion. In contrast, IMU can continuously send its data regardless of environment conditions. Thus, IMU will prove its helpfulness in cases of occlusion by bearing the duty of Kinect in the fusion scheme and correct the false estimation of Kinect during this time.

The following figures illustrate the above scenario:

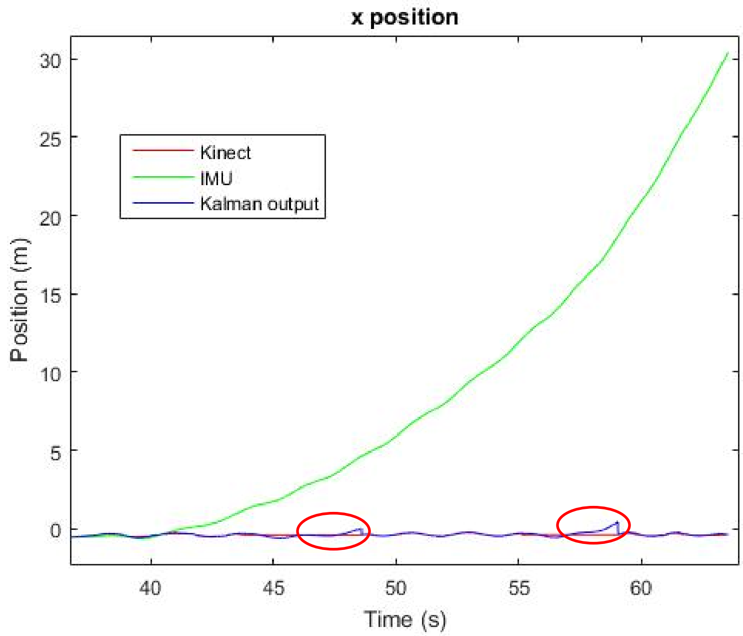

Fig. 10a: Position estimation on x axis by IMU, Kinect and Kalman filter with occlusion

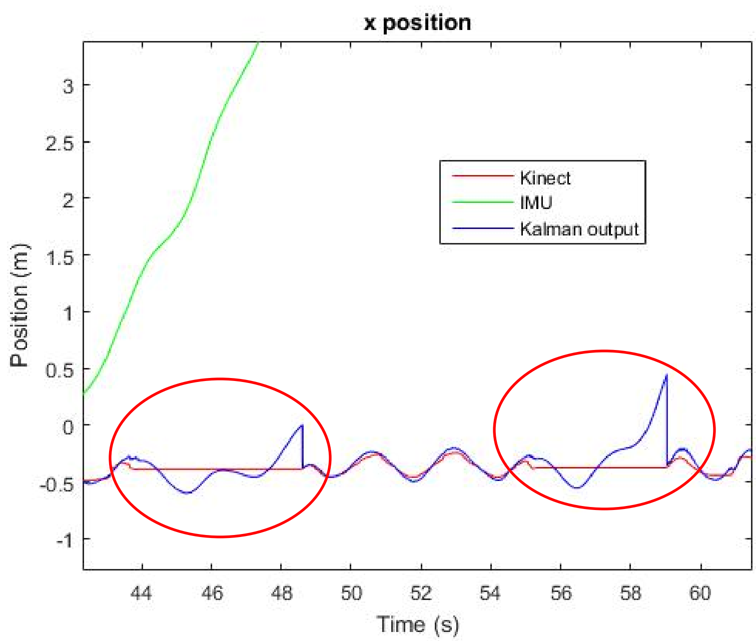

Fig. 10b: Fig. 10a with occlusion intervals zoomed in

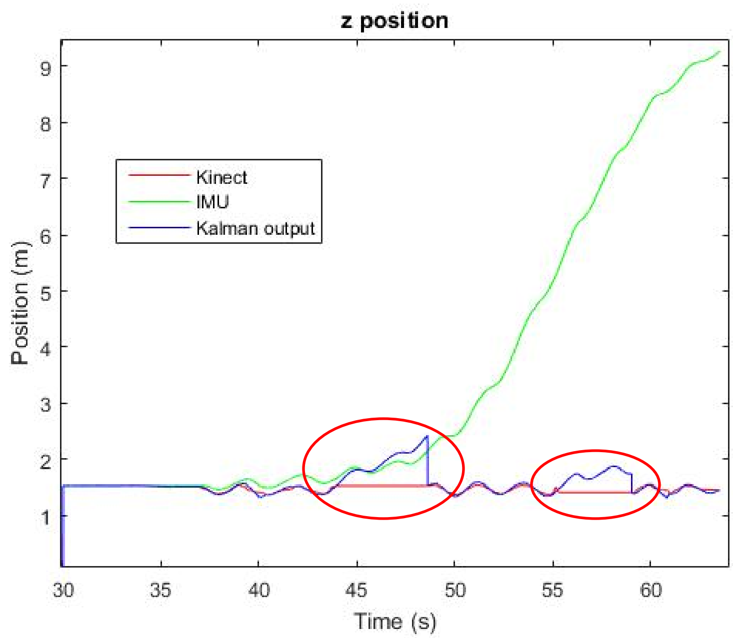

Fig. 10c: Position estimation on z axis by IMU, Kinect and Kalman filter with occlusion

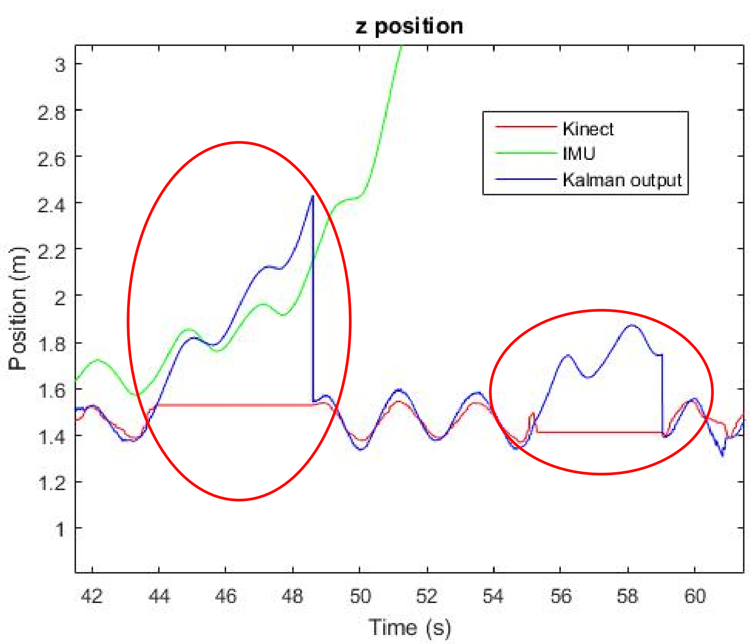

Fig. 10d: Fig. 10c with occlusion intervals zoomed in

It can be seen from Fig. 10 a-d that whenever occlusion happens, Kalman filter will depend heavily on IMU rather than Kinect to estimate position. The fusion thus helps improve accuracy of position estimation by having two modalities helping each other in times of need.

CHAPTER VI

CONCLUSION AND FUTURE WORKS

In this thesis we presented a sensor fusion scheme integrating IMU and Kinect camera and specifically tailored for the application of hand trajectory reconstruction. The results suggested that Kinect, as immuned to drift error, fared better than IMU in terms of position accuracy whenever there is no occlusion. It was also shown that if occlusion did happen in a short period of time, IMU was a great source of help by correcting false estimates of Kinect. It would therefore be recommended that the Kinect covariance be deliberately set much lower than that of IMU in instances of no occlusion, so that Kalman filter would favor the more accurate estimation of Kinect to reconstruct hand trajectory. Conversely, Kinect covariance should be deliberately set much higher than that of IMU in instances of occlusion so that IMU will become the major source for position estimation. By following this recommendation, the proposed fusion scheme successfully overcame each sensor's disadvantages and yielded better estimates than each sensor being used separately.

Future works of this project will include setting up a high accuracy motion capture system (Vicon [5]) to provide ground truth in order to quatitatively evaluate the effectiveness of the current fusion scheme. We will also continue to experiment current Kalman filter algorithm on more complicated shapes, as well as apply more rigorous techniques as we transition to specifically apply the algorithm for handwriting reconstruction.

REFERENCES

[1] Q. Wang, G. Kurillo, F. Ofli, R. Bajcsy, "Evaluation of pose tracking accuracy in the first and second generations of Microsoft Kinect", *Proceedings of IEEE International Conference of Healthcare Informatics*, 2015.

[2] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, and P. Olivier. Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In UIST 2012.

[3] [Online]. Available: http://www.immersion.com.

[4] K. Perng, B. Fisher, S. Hollar, and K. S. J. Pister, “Acceleration sensing glove (ASG),” in *Proc. IEEE Int. Symp. Wearable Comput.*, 1999, pp. 178–180.

[5] [Online]. Available: http://www.vicon.com.

[6] MicroSD Motion Sensor User Manual, ESP Lab internal training document.

[7] J. Gummeson, B. Priyantha, and J. Liu. An energy harvesting wearable ring platform for gesture input on surfaces. In *Mobisys*. ACM, 2014.

[8] I. Frosio, S. Stuani, and N. Borghese, "Autocalibration of mems accelerometer," in Proc. IEEE Instrumentation and Measurement Technology Conf., pp. 519 -523, Apr. 2006.

[9] V. Frati and D. Prattichizzo, "Using Kinect for hand tracking and rendering in wearable haptics," in 2011 IEEE World Haptics Conference (WHC), 2011, pp. 317-321.

[10] Y. Tian, X. Meng, D. Tao, D. Liu, and C. Feng, “Upper limb motion tracking with the integration of IMU and kinect,” *Neurocomputing*, vol. 159, pp. 207–218, Jul. 2015.

[11] J. S. Wang, Y. L. Hsu, J. N. Liu, "An inertial-measurement-unit-based pen with a trajectory reconstruction algorithm and its applications", *IEEE Trans. Ind. Electron*., vol. 57, no. 10, pp. 3508-3521, Oct. 2010.

[12] G. Zhang, G. Shi, Y. Luo, H. Wong, W. J. Li, P. H. W. Leong, and M. Y. Wong, “Towards an ubiquitous wireless digital writing instrument using MEMS motion sensing technology,” in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, 2005, pp. 795–800.

[13] W. C. Bang, W. Chang, K. H. Kang, E. S. Choi, A. Potanin, and D. Y. Kim, “Self-contained spatial input device for wearable computers,” in *Proc. IEEE Int. Conf. Wearable Comput.*, 2003, pp. 26–34.

[14] Destelle F, Ahmadi A, O’Connor NE, Moran K, Chatzitofis A, Zarpalas D, Daras P (2014) Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors. In: Signal Processing Conference (EUSIPCO), 2014 *Proceedings of the 22nd European*, pp 371–375

[15] M. Gabel1, R. Gilad-Bachrach, E. Renshaw, A. Schuster, "Full Body Gait Analysis with Kinect", *Proc Int. Conf. of the IEEE Eng. Med. Biol. Soc.*, 2012-Aug.

[16] Z. Ren, J. Meng, J. Yuan, Z. Zhang, "Robust hand gesture recognition with Kinect sensor", *Proc. ACM Int. Conf. Multimedia*, pp. 759-760, 2011.

[17] Gabel M, Gilad-Bachrach R, Renshaw E, Schuster A: Full body gait analysis with Kinect. In Engineering in Medicine and Biology Society (EMBC), 2012 *Annual International Conference of the IEEE. IEEE*; 2012:1964-1967.

[18] [Online]. Available: https://www.microsoftstore.com/store/msusa/en_US/pdp/Kinect-Sensor-for-Xbox-One/productID.2267482500.

[19] Madgwick SO, Harrison AJ, Vaidyanathan R. Rehabilitation Robotics (ICORR), 2011 *IEEE International Conference on. IEEE;* 2011. Estimation of imu and marg orientation using a gradient descent algorithm; pp. 1–7. [PubMed]

[20] Jian Wu and Roozbeh Jafari, “Zero-Effort Camera-Assisted Calibration Techniques for Wearable Motion Sensors,” in *Proceedings of the Wireless Health* 2014 on National Institutes of Health (WH '14), ACM, 2014.