

Upper limb motion tracking with the integration of IMU and Kinect

Yushuang Tian^a, Xiaoli Meng^b, Dapeng Tao^{c,d,*}, Dongquan Liu^a, Chen Feng^a

^a Singapore Polytechnic, Singapore, Singapore

^b Institute for Infocomm Research, Singapore, Singapore

^c Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

^d The Chinese University of Hong Kong, Hong Kong, China



ARTICLE INFO

Article history:

Received 9 November 2014

Received in revised form

22 January 2015

Accepted 27 January 2015

Communicated by Yue Gao

Available online 19 February 2015

Keywords:

Motion tracking

IMU

Kinect

Sensor fusion

ABSTRACT

Upper limb motion tracking attracts attentions from both academia and industry due to its value in a wide range of applications. Although existing optical-based tracking techniques can provide accurate tracking results, the product cost and complexity keep them away from most daily life applications. Recently, low-cost inertial measurement unit (IMU) and Kinect techniques provide a feasible/economical solution for such trajectory tracking problems while either of them still has its own limitations. In this paper, we investigated how to fuse data from internal sensors of IMU, and fuse IMU data with Kinect in order to provide robust hand position information compensated for the limitations of those sensors. The calculation of position is sequentially achieved by three fusion strategies: double integration of IMU internal sensors, IMU internal sensor fusion with geometrical constraints and unscented Kalman filter (UKF) based fusion of IMU and Kinect. Experimental results show that the first two approaches suffer from drifting effects, while the proposed IMU and Kinect fusion method can provide drift-free and smooth results. Comparing with using Kinect alone, this approach is able to achieve better results in terms of both accuracy as well as robustness.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Human upper limb trajectory/motion tracking is a rapidly growing research topic for human motion tracking [1]. It has broad applications, including gaming, human computer interaction and medical rehabilitation [2–6]. Among motion tracking for human upper limb, 3D position tracking is more complex when compared with angle tracking, such as elbow angle motion tracking, due to its higher degrees of freedom (DOF). In this paper, we focus on human upper limb position tracking techniques and its application in stroke rehabilitation, which is motivated by the increased number of stroke patients [7] over recent years.

Till present, one of the most established techniques is optical-based tracking, such as Qualisys system [8]. In optical tracking systems, small reflective markers, which can be detected during movements, are attached to the human subject's upper limbs. Therefore, the tracking systems can provide accurate tracking results while reduce uncertainty of a subject's upper limb motion. However, an optical-based tracking system needs multiple high-end cameras which should be structured and calibrated within a controlled environment. The environment should be generally big,

thus, it is difficult to set up out of dedicated labs. It also requires a complex system to process a huge amount of data, and this restricts the systems from real-time applications. Moreover, the cost of the products is unaffordable for most daily life applications. Therefore, different alternative approaches are desired, which should provide comparable accuracy but with simpler setup and lower cost.

With the rapid advances of micro-electro-mechanical systems (MEMS) sensors, inertial measurement unit (IMU) which consists of accelerometers, gyroscopes and magnetometers, has been frequently used in human upper limb motion tracking [9–13]. The underlying idea of these methods is based on integration/double integration of inertial sensor data. In [9], a joint angle measurement method was developed based on a pair of virtual sensors located at two adjacent segments. Orientation of each segment is calculated using combination of inertial accelerometers and gyroscopes. And joint angle is obtained from the orientation difference of two adjacent segments. Following this work, Hyde et al. [10] proposed an upper-limb orientation estimation approach which focused on optimizing the system by minimizing sensor number. Some work centered on developing more sophisticated fusion techniques for real-time applications. In [11], Yun et al. proposed an extended Kalman filter to estimate orientation of human limb using combination of inertial sensors and magnetic sensors. Xsens [12] and InterSense [13] had similar work using hybrid inertial sensors and magnetic or ultrasonic sensors. The

* Corresponding author at: Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China.

E-mail address: dapeng.tao@gmail.com (D. Tao).

above methods can be considered as orientation-based motion tracking approaches. Other researchers focused on tracking human subjects' position directly with the help of additional constraints, such as geometrical model [14–16] and zero-velocity update (ZUPT) [17]. Tao et al. [14] presented a novel sensing and data fusion system to track 3D hand motion in a telerehabilitation program. In their work, data from IMU and webcam are fused to estimate the hand trajectories. A physical arm constraint is applied to further improve the system's performance. In [15], Kim et al. proposed a wearable upper limb motion tracking method for stroke rehabilitation therapy at home. Their system consists of two IMU mounted on wrist and elbow. The upper limb motion trajectory is calculated by tracking the position of forearm and upper arm. Subsequently, geometrical models are used to eliminate the position drift errors to improve system accuracy. However, those aforementioned approaches suffer from inherent bias errors existing in inertial data, which will be accumulated with respect to time and decrease the tracking accuracy.

Recently, Kinect sensor is developed by Microsoft Inc. to track 3D joint positions of human subjects, which has been applied in the Xbox series home game system. In [18], Chang et al. compared the motion tracking performance between the low-cost Microsoft Kinect and the high fidelity OptiTrack optical system and show that Kinect can achieve drift-free motion tracking. However, this obtained motion trajectory suffers from position fluctuation as estimated positions by Kinect are sensitive to noise and occlusions. Different from Kinect, IMU can produce smooth motion position tracking results with low accuracy. Therefore, a feasible approach to obtain an accurate and robust tracking result should be based on the compensation of these two sensors. Inspired by this idea, Lanari et al. [19] developed a knee angle estimation method for rehabilitation applications by fusing data from IMU and Kinect together. However, this work focuses on angle estimation, specific geometrical model should be taken into account when extending IMU/Kinect fusion to hand position tracking applications [20,21]. In view of this, we proposed a new IMU and Kinect fusion approach to estimate 3D position of upper limb directly here.

In this paper, first, we investigate upper limb motion tracking techniques with internal sensors of IMU fusion alone and then develop a specific geometrical model to constrain the fusion results. Although such tracking techniques with internal sensors of IMU fusion alone can produce smooth tracking results, the

accuracy is unacceptable for upper limb motion tracking due to the drifting problems. Finally, a sensor fusion algorithm with integration of IMU and Kinect is proposed. This fusion method combines the advantage of IMU sensors and Kinect together to provide accurate and robust upper limb trajectory tracking. As opposed to the fusion approaches with IMU sensor alone, the proposed method integrates Kinect to compensate the bias error of IMU sensors and eliminate the drifting effect in final results. Comparing with tracking results from Kinect alone, the proposed method can reduce position fluctuation and alleviate occlusion sensitivity. Furthermore, a specific geometrical model is developed to restrict the motion estimation. Because of the nonlinearity of dynamic model and measurement model, an UKF is utilized to fuse IMU and Kinect sensor data recursively to estimate the 3D upper limb position.

The rest of the paper is organised as follows. Section 2 provides the configuration of the proposed system. Section 3 presents a customized conventional IMU internal sensor fusion algorithm for position estimation. Physical geometrical constraint is introduced in Section 4 to improve IMU fusion accuracy. In Section 5, the proposed IMU/Kinect fusion algorithm is introduced. Experimental results of all three approaches are compared with the ground truth in Section 6. Finally, a brief conclusion and some extensions for future work are given in Section 7.

2. System and user study configuration

The purpose of this study is to track a subject's upper limb trajectories, which has many real-life applications including medical rehabilitation and home entertainment. Taking the medical rehabilitation as an example, upper limb trajectory of a stroke patient can be recorded by computer. Afterwards, an experienced doctor can evaluate to what extent the patient has recovered based on a number of criterion, such as smoothness of the motion trajectory. In this study, two IMUs are mounted on the subject's right forearm and right upper arm, as shown in Fig. 1. They will record the acceleration $\mathbf{a}_B^{f,u}$ and angular velocity $\omega_B^{f,u}$ in the body frame of the moving hand, where the superscript f and u represent forearm and upper arm, respectively. The Kinect sensor is located in front of the subject, which can provide unbiased position of the subject's upper limb motion trajectory.

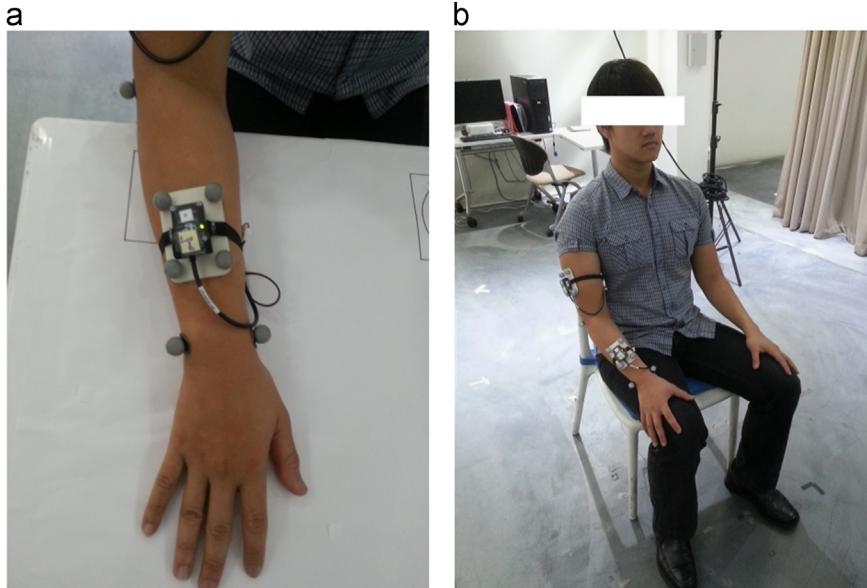


Fig. 1. Placement of IMU. (a) IMU surrounded by four markers; (b) a subject with one IMU placed on forearm and the other IMU on upper arm.

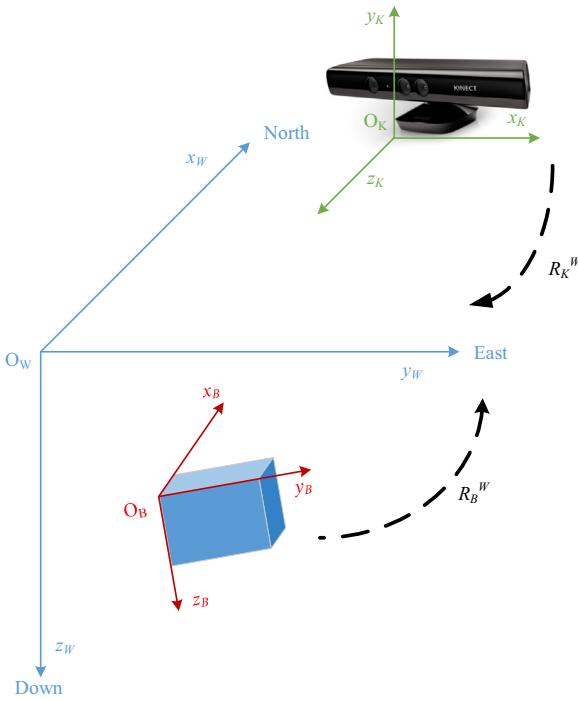


Fig. 2. The body frame $Ox_B y_B z_B$, the global frame $Ox_W y_W z_W$, and the Kinect frame $Ox_K y_K z_K$.

The three coordinate systems involved in the hand trajectory tracking system are shown in Fig. 2, which include (1) the body frame ($Ox_B y_B z_B$); (2) the Kinect frame ($Ox_K y_K z_K$); (3) the global reference frame ($Ox_W y_W z_W$) with X-axis pointing to the North, Y-axis pointing to the East and Z-axis pointing to the Down direction, respectively. R_B^W represents the rotation transformation between the body frame $Ox_B y_B z_B$ and the global frame $Ox_W y_W z_W$. R_K^W is the rotation transformation between the Kinect frame $Ox_K y_K z_K$ and the global frame $Ox_W y_W z_W$. Generally, the position of the subject's hand is calculated in the global frame $Ox_W y_W z_W$. Therefore, the body frame $Ox_B y_B z_B$ and the Kinect frame $Ox_K y_K z_K$ should be projected onto the global frame $Ox_W y_W z_W$ using the rotation transformations R_B^W and R_K^W , respectively.

In our experiments, the subject is required to be seated on a chair with his back against backrest and hand on his/her lap. The activity is to lift one hand, place the hand on mouth and bring it back to the initial position on the lap, which will be repeated for 3 times. The entire motion should be conducted as fast and smooth as possible. This designed task is recommended by therapists based on stroke rehabilitation measurement ARAT (Action Research Arm Test) [22]. For this task, upper limb motion trajectory tracking plays an important role in such stroke rehabilitation applications [23,24] where therapists and doctors could evaluate patient performance based on the obtained motion data. For accuracy comparison and evaluation, optical system Qualysis is used to provide the ground truth of 3D positions. In this study, each IMU is surrounded by four markers, as shown in Fig. 1(a). During experiments, markers' trajectories are recorded by Qualysis. Based on the relative positions between IMUs and markers, actual 3D positions of the IMUs can be calculated by the Qualysis.

3. IMU internal sensor fusion

In this study, Inertiacube 3 IMU [13] from interSense is used. And the underlying principle is based on the strapdown [26] IMU navigation algorithm. In detail, we only utilize the raw data of

angular velocity, acceleration and magnetometer in body frame in the proposed method. The first step is to integrate the angular velocity with respect to time to calculate the orientation. Then, the obtained orientation information is used to calculate the motion acceleration in the global frame. Finally, 3D position is obtained by the double integration of motion acceleration in the global frame.

3.1. Unscented Kalman filter

The data fusion problem in our system is a typically nonlinear problem. It is noted that the well-known Kalman Filter technique is only suitable for linear systems. The Extended Kalman Filter (EKF) has become a standard formulation for nonlinear state estimation. However, it may cause significant error for highly nonlinear systems due to the propagation of uncertainty through the nonlinear system. Recently, the UKF [25] is a novel technique developed in the field. The underlying idea is to produce several sampling points (Sigma points) around the current state estimated based on its covariance. Then, these points are propagated through a nonlinear map to get more accurate estimation of the mean and covariance of the mapping results. Thus, it avoids the need to calculate the Jacobian matrix, hence incurs only the similar computation load when compared with EKF. Therefore, our sensor fusion framework is based on the UKF technique.

The state vector \mathbf{x} consists of the following three parameters: (1) position \mathbf{p} , (2) velocity \mathbf{v} , (3) orientation quaternion \mathbf{q} . It can be expressed as

$$\mathbf{x} = \begin{pmatrix} \mathbf{p} \\ \mathbf{v} \\ \mathbf{q} \end{pmatrix} \quad (1)$$

where $\mathbf{p} = [p_{x_w}, p_{y_w}, p_{z_w}]$ and $\mathbf{v} = [v_{x_w}, v_{y_w}, v_{z_w}]$ are the tri-axis instantaneous position and velocity in the global frame, respectively. $\mathbf{q} = [q_0, q_1, q_2, q_3]$ represents the rotation of the sensor body frame with respect to the global frame.

3.2. Process model

A process model describes the dynamic relationship between two successive states. Generally, the process model can be given by

$$\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t, \mathbf{w}_t), \quad (2)$$

where \mathbf{x}_{t+1} and \mathbf{x}_t represent the state vectors of two successive time step and \mathbf{w}_t is the process noise which drives the dynamic system. The dynamic system \mathbf{F} is a nonlinear function with respect to \mathbf{x} and \mathbf{w} . In what follows, the dynamic system \mathbf{F} will be derived.

First, it is noted that quaternion \mathbf{q} satisfies the following differential equation [16,27,29]:

$$\frac{d}{dt}\mathbf{q} = \frac{1}{2}\Omega[\omega]\mathbf{q} = \frac{1}{2}\begin{pmatrix} -[\omega \times] & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{pmatrix}\mathbf{q} \quad (3)$$

where $\Omega[\omega]$ is a 4×4 skew symmetric matrix, $\boldsymbol{\omega} = [\omega_{x_b}, \omega_{y_b}, \omega_{z_b}]$ is the angular velocity of the body frame with respect to the global frame and $[\omega \times]$ represents the cross product operator as follows,

$$[\omega \times] = \begin{pmatrix} 0 & -\omega_{z_b} & \omega_{y_b} \\ \omega_{z_b} & 0 & -\omega_{x_b} \\ -\omega_{y_b} & \omega_{x_b} & 0 \end{pmatrix} \quad (4)$$

Under the assumption that the angular velocity $\boldsymbol{\omega}_t$ is constant during the time interval Δt , Eq. (3) can be rewritten in the discrete-time form,

$$\mathbf{q}_t = \exp\left(\frac{1}{2}\Omega[\omega_t]\Delta t\right)$$

$$\mathbf{q}_{t-1} = \mathbf{q}_{t-1} \otimes \begin{bmatrix} \frac{\omega_t}{|\omega_t|} \sin\left(\frac{|\omega_t|}{2}\Delta t\right) \\ \cos\left(\frac{|\omega_t|}{2}\Delta t\right) \end{bmatrix} \quad (5)$$

where \otimes represents the quaternion multiplication. The first-order approximation in (5) is obtained following the idea in [29].

Then, the motion acceleration \mathbf{a}_t can be calculated from the inertial acceleration measurement \mathbf{a}_t^B using

$$\mathbf{a}_t = \mathbf{q}_t \otimes \mathbf{a} \otimes \mathbf{q}_t^{-1} - \mathbf{g}, \quad (6)$$

where \mathbf{q}_t^{-1} is the inverse quaternion of \mathbf{q}_t . Pure vector quaternion $\mathbf{a} = (0, \mathbf{a}_t^B)$ and $\mathbf{g} = (0, \vec{\mathbf{g}})$, where $\vec{\mathbf{g}}$ represents the earth gravity.

Subsequently, the velocity and position can be obtained after integration/double integration of the acceleration, which can be rewritten in a discrete-time form as

$$\begin{pmatrix} \mathbf{p}_t \\ \mathbf{v}_t \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \Delta t \mathbf{I} \\ 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{p}_{t-1} \\ \mathbf{v}_{t-1} \end{pmatrix} + \begin{pmatrix} 0.5 \Delta t^2 \mathbf{I} \\ \Delta t \mathbf{I} \end{pmatrix} \mathbf{a}_{t-1}, \quad (7)$$

where $\Delta t = 0.03$ s is the sensor sampling time interval. \mathbf{I} is a 3×3 identity matrix and \mathbf{a}_{t-1} is the acceleration calculated by (6). Beginning with an initial estimation of the process state, the real-time position \mathbf{p}_t can be iteratively calculated based on (7).

3.3. Measurement model

The measurement model serves to relate the measurement vector to the statement vector. The general form of measurement model can be written as

$$\mathbf{y}_t = \mathbf{H}(\mathbf{x}_t, \mathbf{n}_t), \quad (8)$$

where \mathbf{y}_t represents the measurement vectors and \mathbf{n}_t is the measurement noise. The measurement model \mathbf{H} is a function with

respect to \mathbf{y} and \mathbf{n} . The measurement vector \mathbf{y}_t consists of the acceleration in the body frame \mathbf{a}_t^B and the magnetic strength in the body frame \mathbf{m}_t^B as

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{a}_t^B \\ \mathbf{m}_t^B \end{pmatrix}, \quad (9)$$

Let $\vec{\mathbf{m}}$ be the vector of the magnetic field (“north”) in the global frame. The expected measurements of the global frame are given by the transformation of $\vec{\mathbf{g}}$ and $\vec{\mathbf{m}}$ to the body frame, which can be written as

$$\mathbf{a}_t^B = \mathbf{q}_t^{-1} \otimes \mathbf{g} \otimes \mathbf{q}_t + \mathbf{n}_{acc,t}, \quad (10)$$

$$\mathbf{m}_t^B = \mathbf{q}_t^{-1} \otimes \mathbf{m} \otimes \mathbf{q}_t + \mathbf{n}_{mag,t}, \quad (11)$$

where $\mathbf{m} = (0, \vec{\mathbf{m}})$ represent the pure vector quaternions of $\vec{\mathbf{m}}$. $\mathbf{n}_{acc,t}$ and $\mathbf{n}_{mag,t}$ are the measurement noise of acceleration and magnetic strength in the body frame, respectively. After Eqs. (2) and (8) is defined, Matlab UKF toolbox [28] is used to solve this fusion problem and calculate the position estimation.

3.4. Bias estimation

It is noted that the raw data \mathbf{a}_t^B , ω_t^B and \mathbf{m}_t^B contain inherit bias due to thermal noise of the sensors, which will decrease the accuracy of position estimation. Therefore, the bias error should be estimated and removed from the input data. First, we keep the IMU static and record its output data for about 45 min. And then, a fitted error model (12) is applied to the recorded data to describe the acceleration error $\epsilon_{a,t}$ with respect to time t . Finally, the acceleration error $\epsilon_{a,t}$ is removed from the acceleration \mathbf{a}_t^B directly

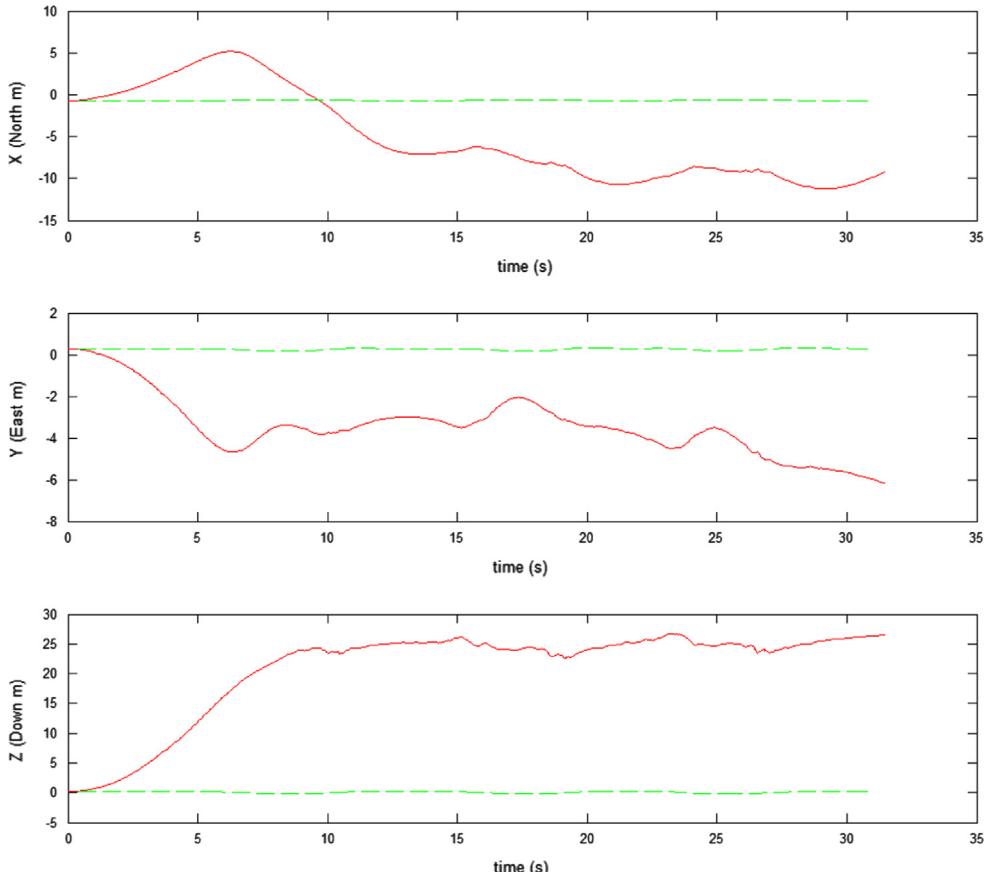


Fig. 3. Forearm trajectories by IMU internal sensor fusion alone with bias correction. Green dashed line: ground truth from Qualysis; red solid line: estimated position calculated based on IMU. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[29,30].

$$e_{a,t} = C_1(1 - e^{-t/T}) + C_2, \quad (12)$$

where C_1 , C_2 and T are parameters which need to be tuned based on the test data. For angular velocity ω_t^B and magnetic strength \mathbf{m}_t^B , the processing are similar.

After double integration of the acceleration, the inherit acceleration error is accumulated over time and increases dramatically with time. Hence, the estimated position will be far away from the real position. It can be seen from the experiment as follows: The IMU is located on the forearm centre of the subject who performs the ARAT “hand to mouth” task. During the experiment, the IMU sensor data are recorded. The above IMU fusion algorithm with error correction of the raw data is performed on the collected data. The trajectories along X, Y and Z axes are shown in Fig. 3, respectively. In the figure, the green dashed line represents the ground truth from Qualysis and red solid line is the estimated position calculated based on IMU. From Fig. 3, it is noted that there are heavily shifting effects in the fusion results from the ground truth, which is unacceptable. Therefore, additional constraints need to be considered, which will be discussed in the next section.

4. IMU fusion with geometrical constraint

In this section, we will utilize additional geometrical constraints to alleviate the drifting effect of the IMU internal sensor fusion in upper limb motion tracking. It is noted that elbow of a normal subject can only have 2DOF: (1) flexion/extension and (2) pronation/supination. In theory, elbow's abduction/adduction is nearly impossible. Therefore, the adduction angle θ can be restricted to an angle close to zero as a geometrical constraint. In this IMU fusion method with geometrical constraint, the desired adduction angle is defined as θ which is between Y-axis of upper arm and the vector \mathbf{V} pointing from upper arm IMU to the forearm IMU, as shown in Fig. 4. Assuming that θ is equal to a very small angle, the Y-axis of the upper arm and the vector \mathbf{V} should be almost orthogonal, which can be represented by dot production as

$$c_t(\mathbf{V}_t^u, \mathbf{V}_t^u) = 0 = \mathbf{V}_t \times \mathbf{Y}_t^u + w_t, \quad (13)$$

where w_t is a Gaussian noise variable. The vector \mathbf{Y}_t^u is the Y-axis of the body frame of the upper arm IMU and \mathbf{V}_t is a vector from the forearm IMU to the upper arm IMU, which are defined in the global frame $Ox_Wy_Wz_W$ as

$$\mathbf{Y}_t^u = \mathbf{q}_t^u \otimes [0, 0, 1, 0] \otimes (\mathbf{q}_t^u)^{-1}$$

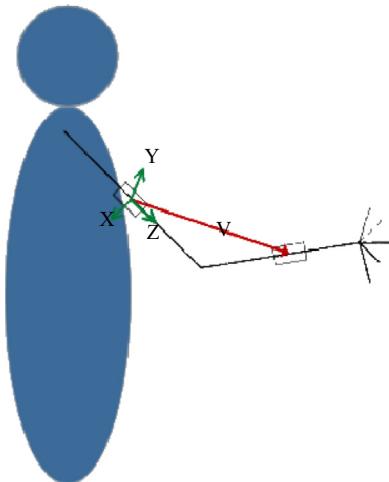


Fig. 4. Illustration of geometrical constraint.

$$\mathbf{V}_t = \mathbf{p}_t^f - \mathbf{p}_t^u \quad (14)$$

where \mathbf{q}_t^u is the orientation of the forearm with respect to the global frame $Ox_Wy_Wz_W$ at time t and it is the quaternion part of forearm state vector \mathbf{x}_t^u . \mathbf{p}_t^f and \mathbf{p}_t^u are the 3D positions of the two IMUs mounted on the upper arm and forearm in the global frame. They are the position parts of forearm state vector \mathbf{x}_t^f and upper arm state vector \mathbf{x}_t^u , respectively.

Now, the upper limb physical geometrical constraint has been established based on Eq. (13). It is noted that this measurement equation is obviously nonlinear. Hence, we again utilize the UKF technique to solve this fusion problem. The dynamical model is represented as

$$\mathbf{x}_t^{c,+} = \mathbf{x}_t^c + \mathbf{r}_t \quad (15)$$

where $\mathbf{x}_t^c = \begin{bmatrix} \mathbf{x}_t^f \\ \mathbf{x}_t^u \end{bmatrix}$, and \mathbf{r}_t is the process noise and is assumed to be a Gaussian vector.

To demonstrate the effectiveness of the geometrical constraint, the elbow adduction angles by two fusion methods are provided in Fig. 5. The result by IMU internal sensor fusion alone is marked in blue and the result by IMU fusion with geometrical constraint is marked in red. It can be seen that the elbow adduction angle obtained by IMU fusion with constraint deviates from 60 degree to 100 degree, which is much smaller than the result obtained by IMU internal sensor fusion alone. The comparison shows that the constraint can keep the adduction movement in a small space effectively. The upper limb motion tracking results obtained by IMU fusion with geometrical constraint are shown in Fig. 6. In the figure, the green dashed line is the ground truth from Qualysis and the red solid line represents the estimated positions based on IMU fusion with constraint. It can be seen that the position can be restricted much closer to the ground truth when compared with the results obtained by IMU fusion alone shown in Fig. 3. However, the motion estimation error is still unacceptable for our rehabilitation application. Therefore, further improvement should be considered to eliminate the drifting effects, which will be explained in the next section.

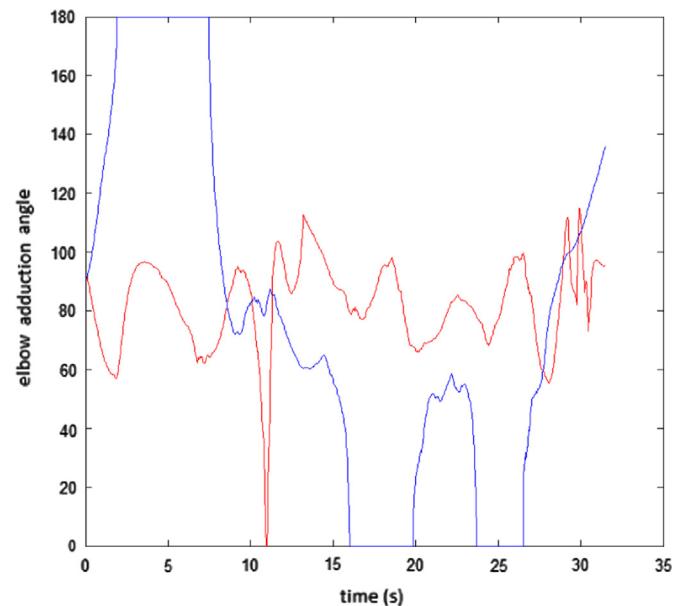


Fig. 5. Comparison of elbow adduction angle by IMU fusion alone (blue) and IMU fusion with geometrical constraint (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

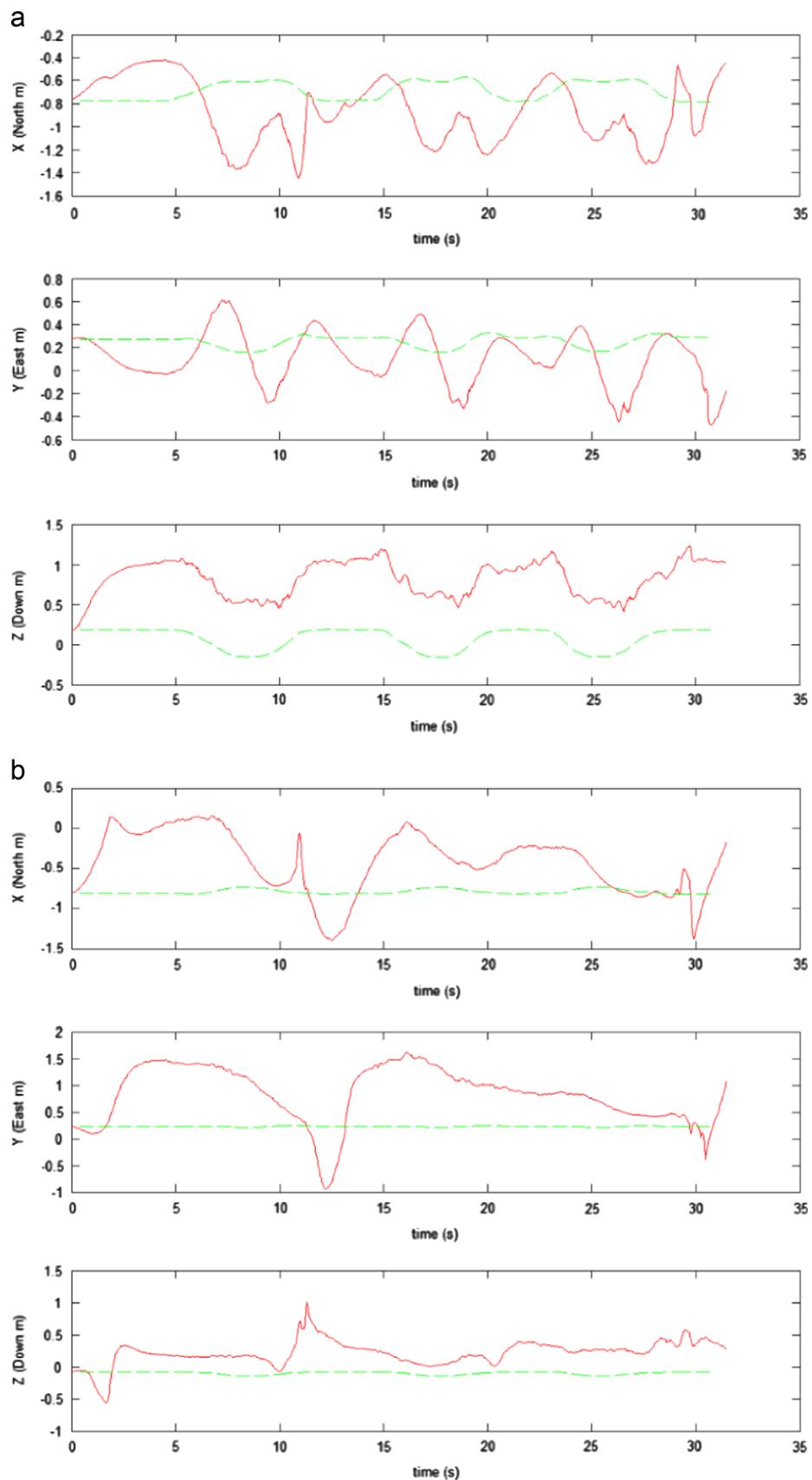


Fig. 6. Upper limb trajectories by the IMU fusion with geometrical constraints. (a) forearm; (b) upper arm. Green dashed line: ground truth from Qualysis; red solid line: estimated position calculated based on IMU. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Sensor fusion of IMU and Kinect

For position estimation, the major disadvantage of IMU fusion algorithms is that it will introduce accumulated drifts over time, which will lead to heavy drifting effects on the positions. In contrast, Kinect can provide positions without drifting error accumulated over time. However, Kinect data are not reliable during motion tracking due to occlusion and joint estimation errors. It is so-called Kinect outage effect which results in jumpy tracking trajectory. It can be observed by the large variance of Kinect positional data to the ground truth from second to second. In this study, such Kinect outage is compensated by IMU, which has the advantage in obtaining smooth trajectories. Therefore, position calculation by combining these two sensors will be superior to either of them alone.

5.1. Transformation between Kinect frame and global frame

It is noted that the position information provided by Kinect cannot be incorporated into previous fusion algorithms directly since the Kinect frame $Ox_Ky_Kz_K$ is not consistent with the global frame $Ox_Wy_Wz_W$, as shown in Fig. 2. A rotation transformation R_K^W is required to align the two coordinate systems. In order to simplify the calculation, we locate the Kinect in front of the subject such that X-axis of the Kinect frame $Ox_Ky_Kz_K$ pointing to the defined East direction, Y-axis pointing to the Up direction and Z-axis pointing to the South direction. From Fig. 2, it can be seen that the axes x_W , y_W and z_W of the global frame are defined in parallel with Kinect frame axes z_K , x_K , and y_K , respectively. Therefore, the relationship between the global frame $Ox_Wy_Wz_W$ and the Kinect frame $Ox_Ky_Kz_K$ can be described as follows

$$\begin{aligned} x_W &= -z_K - z_{\text{shift}} \\ y_W &= x_K - x_{\text{shift}}, \\ z_W &= -y_K - y_{\text{shift}} \end{aligned} \quad (16)$$

where $[x_W, y_W, z_W]$ and $[x_K, y_K, z_K]$ are the coordinates in the global frame $Ox_Wy_Wz_W$ and the Kinect frame $Ox_Ky_Kz_K$, respectively. $[x_{\text{shift}}, y_{\text{shift}}, z_{\text{shift}}]$ represents the translation between the origin of global frame $Ox_Wy_Wz_W$ and the origin of the Kinect frame $Ox_Ky_Kz_K$. In order to measure this translation $[x_{\text{shift}}, y_{\text{shift}}, z_{\text{shift}}]$, we develop a program which is based on depth information provided by Kinect, as shown in Fig. 7 where the global origin is

marked in red “+”. As the tri-axes of the global frame are parallel with the tri-axes of the Kinect frame, we only need to measure the position of global origin in the Kinect frame $Ox_Ky_Kz_K$, which is the translation between the global frame $Ox_Wy_Wz_W$ and the Kinect frame $Ox_Ky_Kz_K$. In our system calibration, the value of the translation vector is $[x_{\text{shift}}, y_{\text{shift}}, z_{\text{shift}}] = [0.002827, -0.104601, 1.615000]$, where all numbers are in meter.

5.2. IMU and Kinect

A high level illustration of proposed Kinect and IMU fusion based upper limb motion tracking algorithm is given in Fig. 8. In summary, there are mainly two components in this algorithm. One is sensor data fusion and the other part is geometrical constraint fusion. At each time instance t , first, sensor data obtained from IMU and Kinect are fused to calculate positions for upper arm and forearm. Then, elbow joint geometrical constraint is used to eliminate drifting effects due to IMU sensor data double integrate. Finally, positional fusion results are updated for the next iteration.

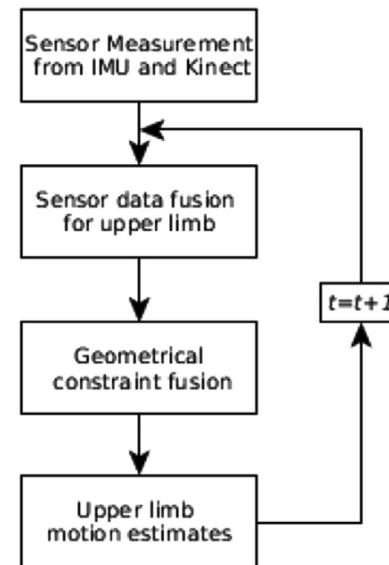


Fig. 8. A high level illustration of upper limb motion tracking.

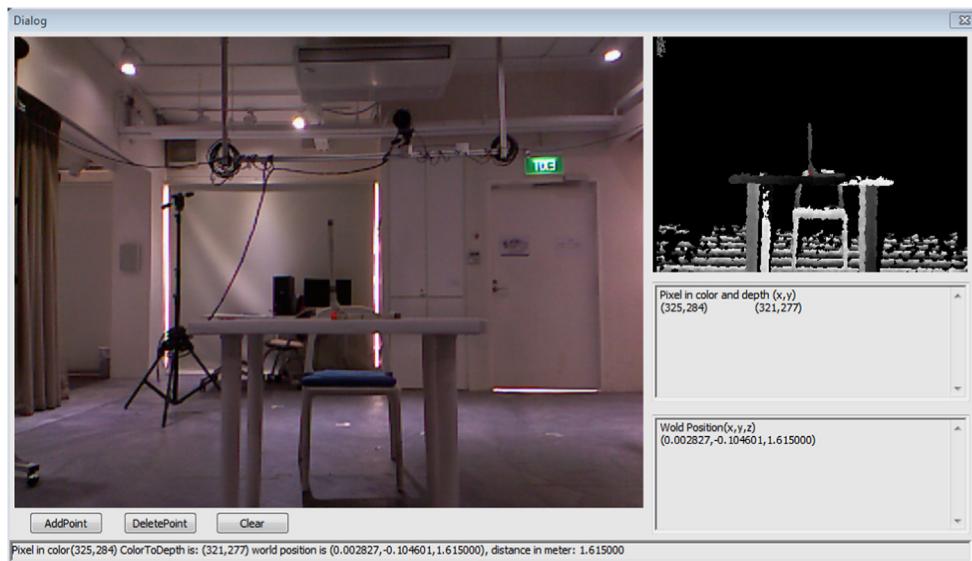


Fig. 7. Measurement of translation between the global frame $Ox_Wy_Wz_W$ and the Kinect frame $Ox_Ky_Kz_K$.

In the sensor fusion processing for upper arm and forearm at time instance t , the state vector \mathbf{x}_t consists of position \mathbf{p}_t , velocity \mathbf{v}_t and orientation quaternion \mathbf{q}_t . As the data fusion problem is a typically nonlinear problem, unscented Kalman filter is selected. Given the state vector \mathbf{x}_t with dimension n and its covariance \mathbf{P}_t , generate $2n+1$ sigma vectors \mathbf{X}_i and weight \mathbf{W}_i , where $i = 0, 1, \dots, 2n$.

$$(\mathbf{X}_0)_t = \mathbf{x}_t$$

$$(\mathbf{X}_i)_t = \mathbf{x}_t + \left(\sqrt{(n+\lambda)\mathbf{P}_t} \right)_i^T, i = 1, \dots, n$$

$$(\mathbf{X}_{i+n})_t = \mathbf{x}_t - \left(\sqrt{(n+\lambda)\mathbf{P}_t} \right)_i^T, i = 1, \dots, n$$

$$W_0^m = \frac{\lambda}{n+\lambda}$$

$$W_0^c = \frac{\lambda}{n+\lambda}$$

$$W_i^m = W_i^c = \frac{1}{2(n+\lambda)}, i = 1, \dots, 2n \quad (17)$$

where $\lambda = \alpha^2(n+k) - n$ is a scaling parameter. α describes the spread of the sigma points around \mathbf{x}_t and is usually a small positive ($\alpha = 10^{-3}$). $k = 0$ is a second scaling parameter. β is used to incorporate prior knowledge of the distribution of \mathbf{x}_t . Here, $\beta = 2$, as \mathbf{x}_t is normally distributed. W_i^m is the weight for the mean associated with the i th sigma point and W_i^c is the weight for the covariance associated with the i th sigma point.

The sensor data fusion for each segment of upper limb is as follows:

- (1) The sigma vectors are propagated through the nonlinear function to yield a set of transformed sigma points based on Eq. (2),

$$(\xi_i)_t = \mathbf{F}((\mathbf{X}_i)_t) \quad (18)$$

- (2) Mean and covariance matrix of $(\xi_i)_t$ is predicted as following

$$\hat{\mathbf{x}}_t^- = \sum_{i=0}^{2n} W_i^m (\xi_i)_t$$

$$\mathbf{P}_t^- = \sum_{i=0}^{2n} W_i^c [(\xi_i)_t - \hat{\mathbf{x}}_t^-] [(\xi_i)_t - \hat{\mathbf{x}}_t^-]^T \quad (19)$$

- (3) The transformed sigma points $(\xi_i)_t$ are propagated through nonlinear function (8),

$$(\mathbf{Y}_i^-)_t = \mathbf{H}((\xi_i)_t) \quad (20)$$

It is noted that this observation model (20) contains position measurement provided by Kinect. In case of Kinect outage, the observation model (20) is equivalent to observation (8).

- (4) Predict observation, measurement covariance matrix and cross covariance matrix,

$$\begin{aligned} \hat{\mathbf{y}}_t^- &= \sum_{i=0}^{2n} W_i^m (\mathbf{Y}_i^-)_t \\ \mathbf{P}_{yy} &= \sum_{i=0}^{2n} W_i^c [(\mathbf{Y}_i^-)_t - \hat{\mathbf{y}}_t^-] [(\mathbf{Y}_i^-)_t - \hat{\mathbf{y}}_t^-]^T \\ \mathbf{P}_{xy} &= \sum_{i=0}^{2n} W_i^c [(\xi_i)_t - \hat{\mathbf{x}}_t^-] [(\mathbf{Y}_i^-)_t - \hat{\mathbf{y}}_t^-]^T \end{aligned} \quad (21)$$

- (5) Updating,

$$\begin{aligned} \mathbf{K}_t &= \mathbf{P}_{xy} \mathbf{P}_{yy}^{-1} \\ \hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_t^- + \mathbf{K}_t (\mathbf{y}_t - \hat{\mathbf{y}}_t^-) \\ \mathbf{P}_t &= \mathbf{P}_t^- - \mathbf{K}_t \mathbf{P}_{yy} \mathbf{K}_t^T \\ \text{where } \mathbf{y}_t &= \begin{pmatrix} \mathbf{p}_t^K \\ \mathbf{a}_t^B \\ \mathbf{m}_t^B \end{pmatrix}. \end{aligned} \quad (22)$$

Generally, the elbow joint geometrical constraint fusion process is similar to the sensor data fusion for upper arm/forearm in Eqs. (18)–(22). In the geometrical constraint fusion process, the state vector is,

$$\mathbf{x}_t^c = \begin{bmatrix} \mathbf{x}_t^f \\ \mathbf{x}_t^u \end{bmatrix} \quad (23)$$

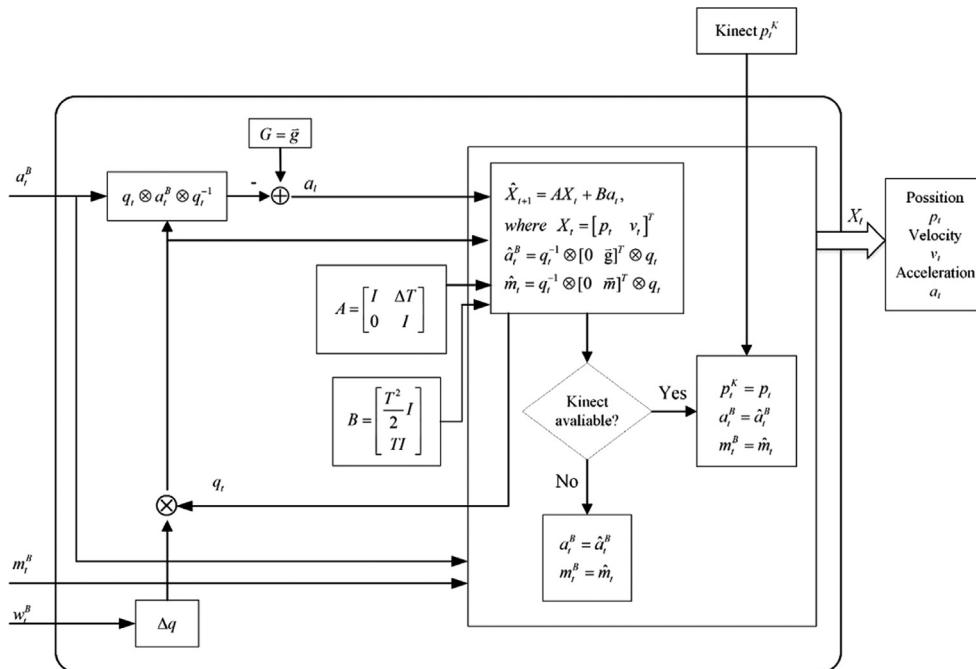


Fig. 9. Flowchart of the IMU and Kinect fusion approach.

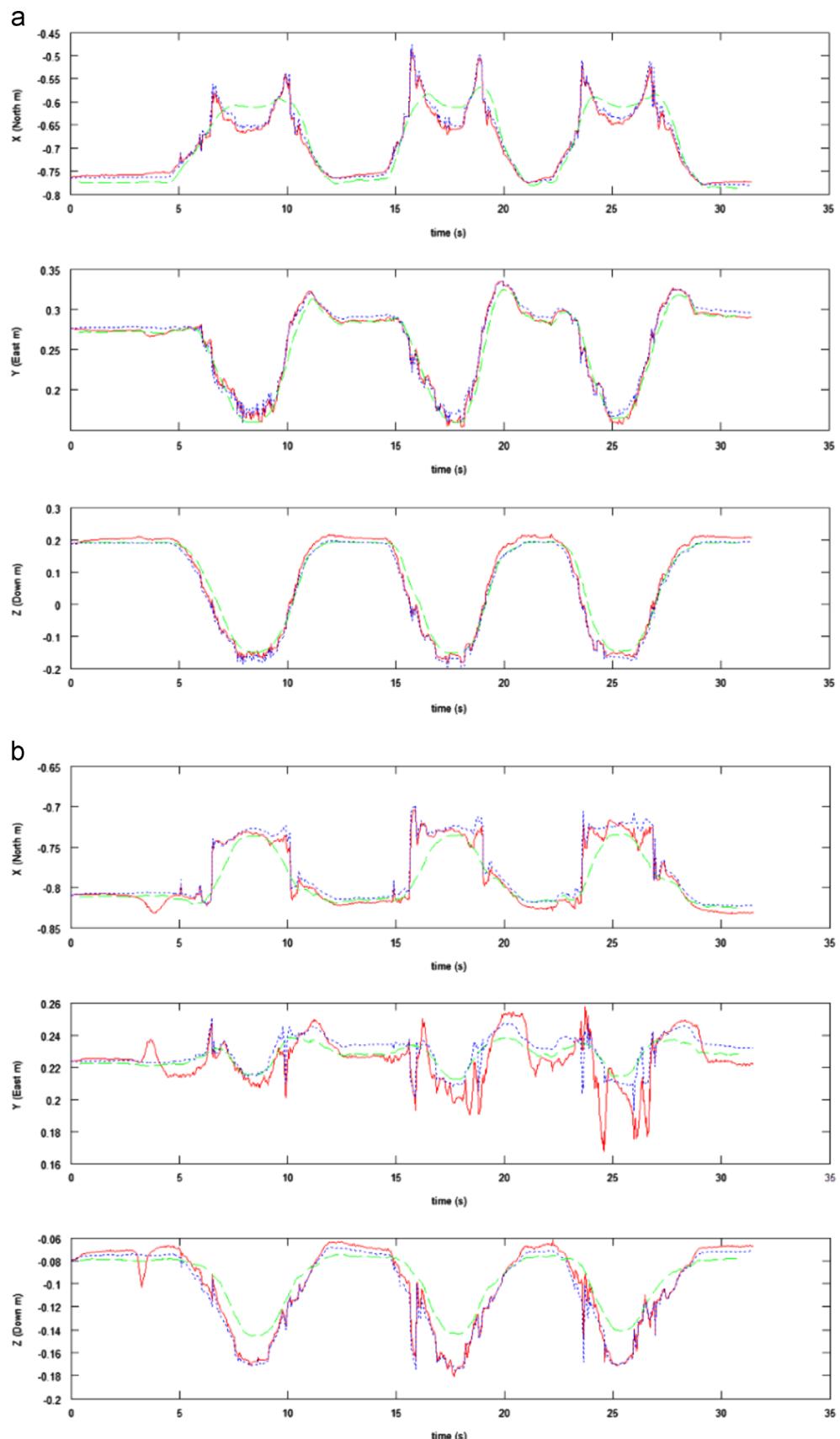


Fig. 10. Fusion results of IMU and Kinect fusion algorithms. Blue dotted line: Kinect result; red solid line: proposed IMU and Kinect fusion algorithm; green dashed line: ground truth from Qualysis. (a) forearm data; (b) upper-arm data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where \mathbf{x}_t^f and \mathbf{x}_t^u are the sensor fusion state vectors of forearm and upper arm, respectively. The construction of process model \mathbf{F} is based on Eq. (15) and construction of measurement model \mathbf{H} is based on Eq. (13).

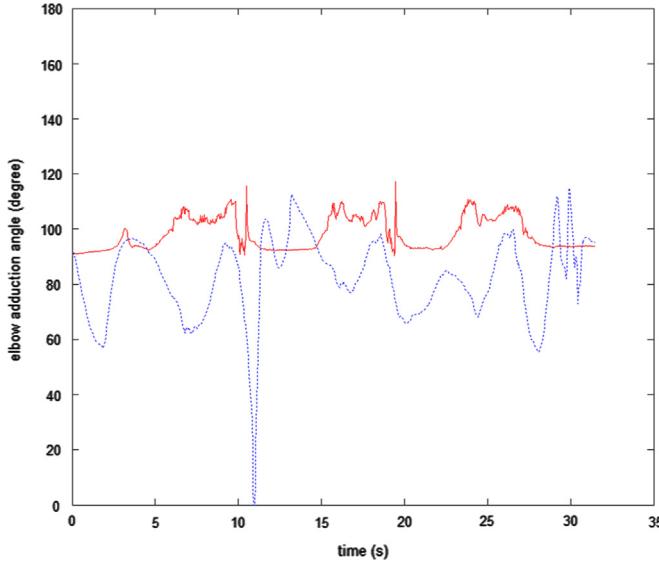


Fig. 11. Comparison of elbow adduction angle by IMU with constraint and IMU and Kinect fusion. Blue dotted line: IMU fusion with constraint; red solid line: proposed IMU and Kinect fusion algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, the sensor data fusion results are updated for the next iteration. Fig. 9 shows above process in flowchart.

6. Experimental results

In this section, various experiments were conducted to illustrate the effectiveness of IMU/Kinect fusion methods. The experiments were conducted on the data collected from a subject performing “hand to mouth” task. The fusion results are given in Fig. 10. It can be observed that IMU and Kinect fusion method provides smoother trajectory (red solid line) when compared with Kinect (blue dotted line). The fusion results from the IMU and Kinect fusion method are much closer to the ground truth obtained by Qualisys (green dashed line) comparing with the results on the same task by IMU fusion with geometrical constraint (shown in Fig. 6). This is because Kinect provides no-drifting measurements during the tracking process. From this comparison, it can be seen that the proposed IMU and Kinect fusion method outperforms both IMU fusion with geometrical constraint and Kinect position tracking approaches. The elbow adduction angle is marked in red, as shown in Fig. 11. It can be seen that angle deviates from the expected value for less than 20 degree. The adduction angle by IMU fusion with constraint is also provided for comparison, which is marked in blue. The comparison results show that the IMU and Kinect fusion method is superior to the method of IMU fusion with constraint and can produce more accurate motion estimation.

In order to further evaluate the robustness of the Kinect and IMU fusion algorithm, another experiment is conducted on

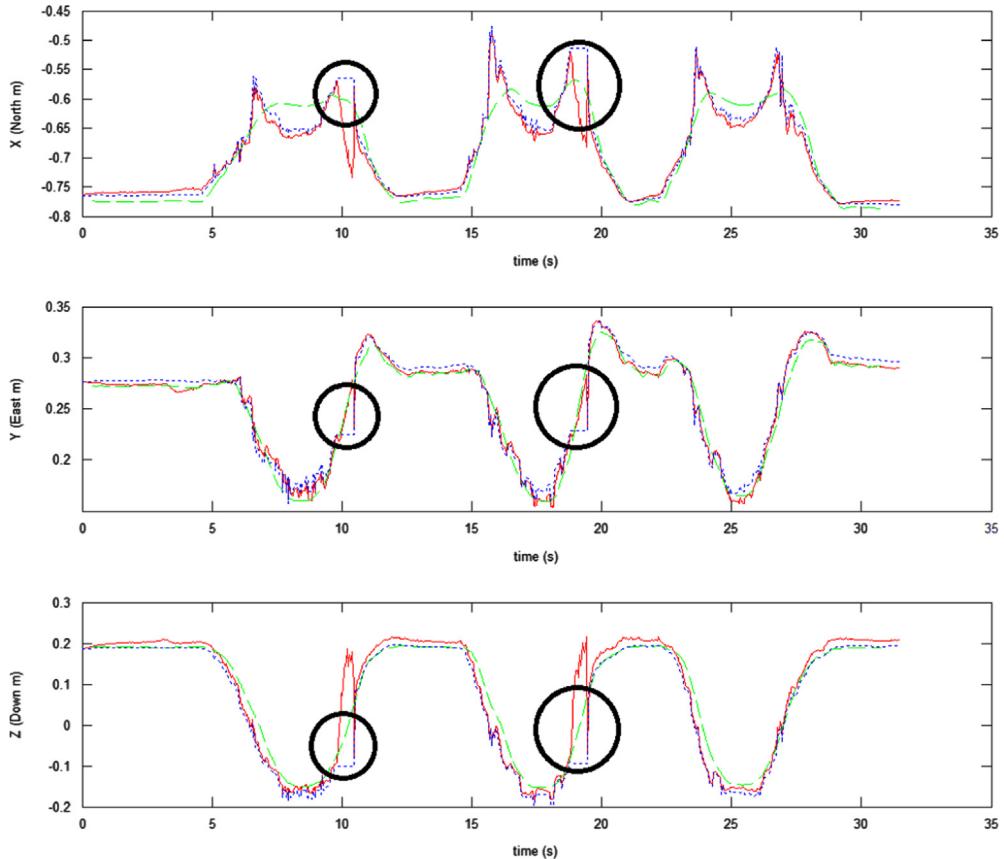


Fig. 12. IMU and Kinect fusion result on forearm data with Kinect outage. Blue dotted line: Kinect result; red solid line: proposed IMU and Kinect fusion algorithm; green dashed line: ground truth from Qualisys; Kinect outages are marked in black circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

forearm data with Kinect outages. The fusion results by various methods are given in Fig. 12. There are two intervals where there are Kinect outages, marked in black circles. From the figure, it is noted that the proposed algorithm can produce almost the same trajectory with when Kinect data are available. This is because the adopted geometrical constraint can restrict the motion estimation back around the expected value when the Kinect data are missing.

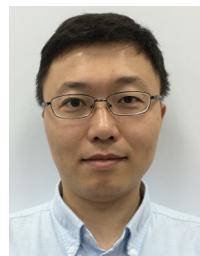
Based on above experimental results, it can be concluded that the proposed IMU and Kinect fusion method is able to effectively and robustly estimate accurate upper limb trajectory by fusing the data from IMU and Kinect sensors.

7. Conclusion

This paper investigated IMU/Kinect fusion techniques for human upper limb trajectory tracking. The double integration of IMU suffers from significant drift problem in position estimation due to existing bias errors in raw data. With emphasizing the geometrical constraint to the motion estimation, the IMU fusion approach with physical geometrical constraint can achieve better results than the double integration IMU fusion method. However, the tracking accuracy is still unacceptable for upper limb motion rehabilitation applications. Comparing with above two approaches, the proposed IMU/Kinect fusion method which utilizes UKF to fuse the data from IMU and Kinect can effectively correct accumulated errors of IMU at the same time overcome the instability of Kinect. Experimental results show that this fused system can provide stable position information with no long-term drifts, which demonstrates an improvement in terms of both accuracy and robustness. Given the low cost of IMU and Kinect comparing with high-end motion capture systems, this work has good potential in real world applications such as home-based rehabilitation. In the future, more effective geometric constraints based on kinematic models will be investigated and experiments on more rehabilitation tasks will be conducted.

References

- [1] H. Zhou, H. Hu, Human motion tracking for rehabilitation—a survey, *Biomed. Signal Process. Control* 3 (1) (2008) 1–18.
- [2] D.S. Nichols-Larsen, P.C. Clark, A. Zeringue, A. Greenspan, S. Blanton, Factors influencing stroke survivors' quality of life during subacute recovery, *Stroke* 36 (2005) 1480–1484.
- [3] M.C. Cirstea, M.F. Levin, Improvement of arm movement patterns and endpoint control depends on type of feedback during practice in stroke survivors, *Neurorehabil. Neural Repair* 21 (2007) 398–411.
- [4] J.A. Gomez-Garcia, J.I. Godino-Llorente, G. Castellanos-Dominguez, Non uniform Embedding based on Relevance Analysis with reduced computational complexity: application to the detection of pathologies from biosignal recordings, *Neurocomputing* 132 (2014) 148–158.
- [5] H. Liang, T. Morie, A motion detection model inspired by hippocampal function and its applications to obstacle detection, *Neurocomputing* 129 (2014) 59–66.
- [6] Y. Choi, S. Ozawa, M. Lee, Incremental two-dimensional kernel principal component analysis, *Neurocomputing* 134 (2014) 280–288.
- [7] [online] <<http://www.strokecenter.org/patients/about-stroke/stroke-statistics/>>.
- [8] [online] <<http://qualisys.se/>>.
- [9] H. Dejnabadi, B.M. Jolles, K. Aminian, A new approach to accurate measurement of uniaxial joint angles based on a combination of accelerometers and gyroscopes, *IEEE Trans. Biomed. Eng.* 52 (8) (2005) 1478–1484, Aug.
- [10] R. Hyde, L. Ketteringham, S. Neild, R. Jones, Estimation of upperlimb orientation based on accelerometer and gyroscope measurements, *IEEE Trans. Biomed. Eng.* 55 (2) (2008) 746–754, Feb.
- [11] X. Yun, E. Bachmann, Design, implementation, and experimental results of a quaternion-based Kalman filter for human body motion tracking, *IEEE Trans. Rob.* 22 (6) (2006) 1216–1227, Dec.
- [12] D. Roetenberg, H. Luinge, P. Slycke, Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors, 2009, pp. 1–7.
- [13] E. Foxlin, M. Harrington, Weartrack: a self-referenced head and hand tracker for wearable computers and portable VR, in: Proceedings of Fourth International Symposium on Wearable Computers, 2002, pp. 155–162.
- [14] Y. Tao, H. Hu, A novel sensing and data fusion system for 3D arm motion tracking in telerehabilitation, *IEEE Trans. Instrum. Meas.* 57 (5) (2008) 1029–1040, May.
- [15] J. Kim, S. Yang, M. Gerla, StrokeTrack: wireless inertial motion tracking of human arms for stroke telerehabilitation, in: Proceedings of the First ACM Workshop on Mobile Systems, Applications, and Services for Healthcare (mHealthSys '11), ACM, New York, NY, USA, 2011.
- [16] Z. Zhang, Z. Huang, J. Wu, Hierarchical information fusion for human upper-limb motion capture, in: Proceedings of 12th International Conference on Information Fusion, 2009, pp. 1704–1711.
- [17] L. Chen, H. Hu, Imu/gps based pedestrian localization, in: Fourth IEEE Computer Science and Electronic Engineering Conference (CEEC), 2012, pp. 23–28.
- [18] C. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A.A. Sawchuk, A.A. Rizzo, Towards pervasive physical rehabilitation using Microsoft Kinect, in: Sixth International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), vol. 162, 21–24 May 2012, 2012, pp. 159.
- [19] A. Padilla, M. Hayashibe, P. Poignet, Joint angle estimation in rehabilitation with inertial sensors and its integration with kinect, in: Annual International Conference of the IEEE EMBS, vol. 8/11, 2011, pp. 3479–3483.
- [20] S. Mefoued, A robust adaptive neural control scheme to drive an actuated orthosis for assistance of knee movements, *Neurocomputing* 140 (2014) 27–40.
- [21] J. Jeong, Normalized estimate of adaptive filter for non-minimum phase transfer function component: an illustrative case study, *Neurocomputing* 133 (2014) 295–308.
- [22] J. Zariffa, N. Kapadia, J. Kramer, P. Taylor, M. Alizadeh-Meghravi, V. Zivanovic, U. Alibisser, R. Willms, A. Townson, A. Curt, Relationship between clinical assessments of function and measurements from an upper-limb robotic rehabilitation device in cervical spinal cord injury, *IEEE Trans. Neural Syst. Rehabil. Eng.* (2011) 341–350.
- [23] L. Dipietro, H. Krebs, B.T. Volpe, J. Stein, C. Bever, S.T. Mernoff, S.E. Fasoli, N. Hogan, Learning, not adaptation, characterizes stroke motor recovery: evidence from kinematic changes induced by robot-assisted therapy in trained and untrained task in the same workspace, *IEEE Trans. Neural Syst. Rehabil. Eng.* 20 (1) (2012) 48–57, Jan.
- [24] Hermano I. Krebs, Michael Kram, Dimitris K. Agrafiotis, Allitia DiBernardo, Juan C. Chavez, Gary S. Littman, Eric Yang, Geert Byttebier, Laura Dipietro, Avrielle Rykman, Kate McArthur, Karim Hajjar, Kennedy R. Lees, Bruce T. Volpe, Robotic measurement of arm movements after stroke establishes biomarkers of motor recovery, *Stroke* 45.1 (2014) 200–204.
- [25] E. Wan, R. van der Merwe, The unscented Kalman filter (chap. 7), in: S. Haykin (Ed.), *Kalman Filtering and Neural Networks*, Wiley, New York, NY, 2001.
- [26] O. Woodman, An Introduction to Inertial Navigation, University of Cambridge, Computer Laboratory, Tech. Rep. UCAMCL-TR-696, 2007.
- [27] D. Choukroun, I.Y. Bar-Itzhack, Y. Oshman, Novel quaternion Kalman filter, *IEEE Trans. Aerosp. Electron. Syst.* 41 (2006) 174–190.
- [28] [online] <<http://becs.aalto.fi/en/research/bayes/ekfukf/>>.
- [29] Z. Zhang, J. Wu, A novel hierarchical information fusion method for three-dimensional upper limb motion estimation, *IEEE Trans. Instrum. Meas.* 60 (11) (2011) 3709–3719, Nov.
- [30] B. Barshan, H.F. Durrant-Whyte, An inertial navigation system for a mobile robot, in: Proceedings of 1st IAV, Southampton, U.K., Apr. 18–21, 1993, 1993, pp. 54–59.



Yushuang Tian received the B.Eng. degree in information engineering from Zhejiang University, Hangzhou, China, in 2005, the M.Sc. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 2007 and the Ph.D. degree in Electrical and Electronic Engineering from Nanyang Technological University, Singapore in 2014. He is currently a research engineer in the School of Mechanical and Aeronautical Engineering at Singapore Polytechnic. His research interests include human motion tracking, computer vision, image/video processing, and image/video super-resolution.



Xiaoli Meng received the B.E. degree in communication engineering from Tianjin University, Tianjin, China, in 2007, and the Ph.D. degree from the Sensor Network and Application Research Center, Graduate University, Chinese Academy of Sciences, Beijing, China, in 2012. She was a Research Fellow with the Department of Bioengineering, National University of Singapore, Singapore, from 2012 to 2013. She is currently a Research Scientist with the Institute for Infocomm Research, Singapore. Her current research interests include autonomous vehicle localization, human motion tracking, gait analysis, and rehabilitation robotics.



Dapeng Tao received a B.Eng. degree from Northwestern Polytechnical University and a Ph.D. degree from South China University of Technology. Over the past years, his research interests include machine learning, computer vision and cloud computing.



Chen Feng received a B.Eng. degree from Southeast University and an M.Sc. degree from Nanyang Technological University. Over the past years, her research interests include medical imaging and Kinect-based applications.



Dongquan Liu received the B.E. degree from Xi'an Jiaotong University, China in 2000 and the Ph.D. degree in Electrical and Electronic Engineering from Nanyang Technological University, Singapore in 2008. He is currently a research scientist in the School of Mechanical and Aeronautical Engineering at Singapore Polytechnic. His research interests include computer vision, machine learning and computer graphics technologies in applications such as computer animation and healthcare.