



Article

Extreme Learning Machine/Finite Impulse Response Filter and Vision Data-Assisted Inertial Navigation System-Based Human Motion Capture

Yuan Xu ¹ , Rui Gao ¹, Ahong Yang ², Kun Liang ¹, Zhongwei Shi ¹, Mingxu Sun ¹ and Tao Shen ^{1,*}

¹ School of Electrical Engineering, University of Jinan, Jinan 250022, China; cse_xuy@ujn.edu.cn (Y.X.); 202221200878@stu.ujn.edu.cn (R.G.); 202130308035@stu.ujn.edu.cn (K.L.); 202130310075@stu.ujn.edu.cn (Z.S.); cse_sunmx@ujn.edu.cn (M.S.)

² School of Music, University of Jinan, Jinan 250022, China; som_yangah@ujn.edu.cn

* Correspondence: cse_st@ujn.edu.cn; Tel.: +86-186-6896-5368

Abstract: To obtain accurate position information, herein, a one-assistant method involving the fusion of extreme learning machine (ELM)/finite impulse response (FIR) filters and vision data is proposed for inertial navigation system (INS)-based human motion capture. In the proposed method, when vision is available, the vision-based human position is considered as input to an FIR filter that accurately outputs the human position. Meanwhile, another FIR filter outputs the human position using INS data. ELM is used to build mapping between the output of the FIR filter and the corresponding error. When vision data are unavailable, FIR is used to provide the human posture and ELM is used to provide its estimation error built in the abovementioned stage. In the right-arm elbow, the proposed method can improve the cumulative distribution functions (CDFs) of the position errors by about 12.71%, which shows the effectiveness of the proposed method.

Keywords: INS; vision; ELM; FIR; human position



Citation: Xu, Y.; Gao, R.; Yang, A.; Liang, K.; Shi, Z.; Sun, M.; Shen, T. Extreme Learning Machine/Finite Impulse Response Filter and Vision Data-Assisted Inertial Navigation System-Based Human Motion Capture. *Micromachines* **2023**, *14*, 2088. <https://doi.org/10.3390/mi14112088>

Academic Editor: Arman Roohi

Received: 2 October 2023

Revised: 29 October 2023

Accepted: 8 November 2023

Published: 12 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, human motion capture has garnered considerable attention, owing to its diverse applications in entertainment, healthcare, and sports industries [1,2]. Accurate motion capture is essential for realistic animation, immersive virtual reality experiences, and a precise biomechanical analysis of human movements [3,4]. Traditional optical motion capture systems are widely employed, but they often exhibit certain limitations, such as high cost, restricted mobility, and dependency on controlled environments [5,6].

Many approaches have been proposed for human motion capture [7]. One of the most famous examples is the vision-based method. For instance, in [8], a vision-based system for tracking and interpreting leg motion in image sequences using a single camera is developed, which is implemented on a commercial computer without any special hardware. A new method for fast human motion capture based on a single red–green–blue–distance (RGB-D) sensor is proposed [9]. To the visual-based human posture capture device, Microsoft's Kinect camera has gained popularity as a depth-sensing device that can capture human movements with high accuracy [10]. This camera utilizes infrared sensors to measure the distance between objects and the camera, generating a detailed three-dimensional point cloud representation of a scene [11]. This depth information, combined with RGB data, enables the precise tracking of human skeletal joints and facilitates real-time motion capture [12]. Depth cameras, integral to Kinect's operation, rely on the emission and detection of infrared light to create depth maps of the surrounding scene [13,14]. However, the presence of environmental obstructions can introduce uncertainties into the captured data. Occurrences where the human body is temporarily occluded by objects within the view of the camera can lead to data gaps or inaccuracies in the motion capture process.

These transient interruptions can impede the seamless reconstruction of motion trajectories, potentially affecting the fidelity and reliability of the captured human movement. In order to deal with this problem, inertial sensors have been proposed to measure human body movements. For instance, in [15], data on human activities are derived from the mobile device's inertial sensor. Meanwhile, Beshara and Chen proposed the use of inertial sensors and Kinect cameras to capture human body movements in their study [16]. It should be noted that although the inertial sensor is able to achieve the seamless measurement, the sensors will experience cumulative errors.

Based on the measurement technology, the data fusion filter will also improve the accuracy of the measurement. To the data fusion filter, it should be pointed out that the Kalman filter (KF) has been widely used. For instance, in [17], the distributed Kalman filter has been proposed to provide a human's position. The dual predictive quaternion Kalman filter is designed for the tracking of the human lower limb posture [18]. Moreover, one new adaptive extended Kalman filter (EKF) for cooperative localization is proposed [19], which is based on the nonlinear system. Moreover, the sigma-point update of cubature Kalman filter is proposed in [20]. One can easily find that the Kalman filter's performance depends on the model's accuracy and the noise description; however, it may be difficult to obtain in practice. In order to overcome this shortcoming, the finite impulse response (FIR) filter is proposed [21]. For example, the extended FIR (EFIR) is used to fuse the inertial navigation system (INS) data and the ultra-wide band (UWB) data. It should be pointed out that the approaches mentioned above do not consider the data outage, which may make the filter's measurement unavailable. In order to overcome this problem, one least-squares support vector machine (LS-SVM)-assisted FIR filter is proposed. In [22], one self-learning square-root cubature Kalman filter is proposed for the integrated global positioning system (GPS)/INS in GPS-denied environments.

To address the limitations of standalone INS and overcome the data gaps in Kinect measurements [23,24], a previous study proposed the use of the extreme learning machine (ELM) algorithm to establish new signals through mapping when UWB signals are interrupted [25]. This allows the entire system to properly function. Building upon this concept, this paper proposes an integrated human motion capture system using ELM, FIR filtering, and INS data assisted by Microsoft's Kinect camera [26], which can reconcile the strengths of INS and Kinect while reducing their weaknesses. The proposed methodology is outlined below.

The INS comprises miniature inertial sensors strategically placed on a subject's body to measure accelerations and angular velocities [27,28]. Raw INS data provide real-time information about the orientation and motion of the subject [29]. The INS comprises miniature inertial sensors accurately placed on the subject's body to measure attitude angles [30]. The raw INS data provide real-time information about the subject's orientation and motion [31] and serve as a foundation for subsequent processes [32]. Meanwhile, the pivotal role of ELM lies in learning the intricate relation between INS-derived body pose data and the corresponding pose data acquired from Kinect [33,34]. Using a shallow neural network architecture, ELM efficiently maps the INS measurements to the corresponding Kinect-based body poses.

Before utilizing ELM, FIR is applied to both the INS data and the pose data obtained from Kinect [17,35]. This filtering process effectively eliminates sensor noise and mitigates the effects of drift, ensuring the accuracy and reliability of the motion capture system [36]. Finally, other researchers previously mentioned that the use of an interactive multiple model (IMM) filtering algorithm can further enhance positioning accuracy. Building upon this idea, the IMM filtering algorithm is employed. The IMM filter is adopted to fuse the INS data and the vision data from Kinect, alongside the ELM-processed data [37]. This fusion process compensates for the missing or erroneous Kinect measurements and further enhances the accuracy of the motion capture system [38,39].

By integrating INS, Kinect vision data, ELM algorithms, IMM algorithms, and FIR filtering, the proposed approach offers an advanced solution for human motion capture.

This integration effectively reduces issues related to data gaps caused by environmental obstructions and high noise levels during Kinect measurements, contributing to improved precision and positioning accuracy in motion capture [40,41]. The resulting system ensures accurate and reliable real-time motion capture, thereby opening up a wide range of possibilities for applications in animation, virtual reality, sports analysis, and healthcare.

To obtain accurate position information, a one-assistant method fusing ELM/FIR filters and vision data is proposed for INS-based human motion capture. In the proposed method, when vision is available, the vision-based human position is inputted into an FIR filter that outputs accurate human position. Meanwhile, another FIR filter outputs the human position using INS data. Moreover, ELM is used to build mapping between the output of FIR and the corresponding error. When vision data are unavailable, FIR is used to provide human posture and ELM is used to provide the estimation error built in the aforementioned stage. Test results confirm the effectiveness of the proposed method.

The main contributions of this study are as follows:

- A seamless INS/vision human motion capture scheme is designed.
- A dual ELM/FIR-integrated filtering is derived.
- An INS/vision human motion capture system is built.
- Experimental evidence shows the better performances of the proposed algorithms than those of traditional algorithms.

The rest of this paper is structured as follows. Section 2 discusses the principle of an INS-based human motion capture system. Section 3 presents the design of an ELM/FIR filter for the human motion capture system. The investigation of experimental tests is summarized in Section 4, and the conclusions are given in Section 5.

2. INS-Based Human Motion Capture System

In this section of the study, the scheme of INS-based human motion capture is designed. The principle of the INS-based human motion capture system is illustrated in Figure 1. In the system used herein, we employ two types of sensors: IMU and vision. First, IMUs are fixed on a target person and employed to measure the posture of the target person's torso. In this work, we employ eight IMUs to measure the acceleration and gyroscope values of the eight key joint points of the target human body. With the measured torso posture and corresponding torso length, the position of the target human's joint points $L_t^{I,j}, j \in [1, 8]$ can be calculated. Meanwhile, we employ the Kinect-based vision sensor to measure the corresponding eight key joint points' vision-derived position $L_t^{V,j}, j \in [1, 8]$.

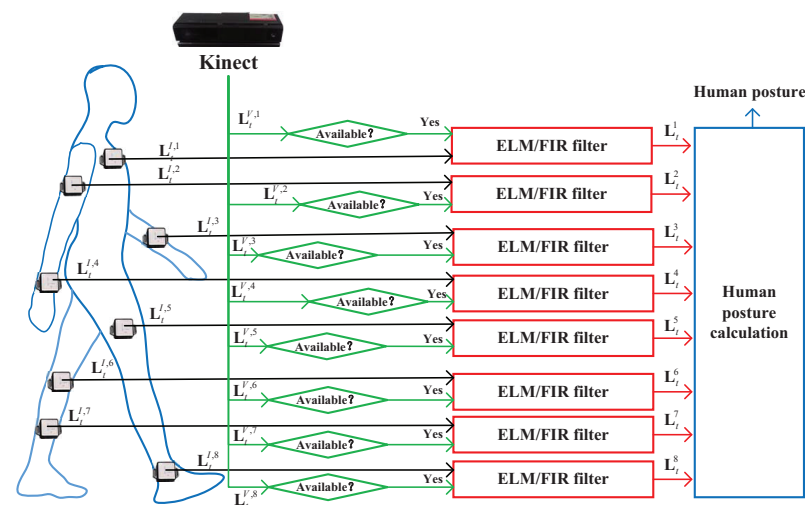


Figure 1. Principle of the INS-based human motion capture system.

From Figure 1, we can see that when vision data are available, the vision-derived position of the target human's joint points $L_t^{V,j}, j \in [1, 8]$ can be obtained. Both $L_t^{I,j}$ and

$\mathbf{L}_t^{V,j}$ are input to the ELM/FIR filter, which is derived in the following section. Then, the ELM/EFIR filter outputs the estimated target human's joints' position $\mathbf{L}_t^j, j \in [1, 8]$, which is used to compute the human posture. Notably, visual data are not continuously provided to the human pose measurement system in practical applications. When vision data are unavailable, only the $\mathbf{L}_t^{L,j}$ are used by the ELM/FIR filter. Note that we only consider the human posture in this work, and the Kinect data may be difficult to obtain, which may result in data outage. It should be emphasized that a Kinect data outage of the eight joints does not occur simultaneously, but randomly. The main motivation of our proposed algorithm in this work is that it can provide relatively accurate pose information when the two types of sensors are working properly. Once visual information experiences an outage, it can also ensure the normal operation of the pose system.

Calculation of the Human Joints' Position

In this section, the calculation of the human joints' position is proposed. First, the coordinate system and key parameters of the human body used herein are introduced. Second, the method of data measurement using the inertial and visual sensors is designed. Figure 2 shows the body frame (b-frame) and the navigation frame (n-frame) as well as the key parameters of the human body used herein. We employ the upper chest's position as the coordinate origin. Only the positions of the elbows, wrist, knee, and ankle ($\mathbf{Po}_2^L, \mathbf{Po}_2^R, \mathbf{Po}_3^L, \mathbf{Po}_3^R, \mathbf{Po}_4^L, \mathbf{Po}_4^R, \mathbf{Po}_5^L$, and $\mathbf{Po}_5^R$ shown in Figure 2) are considered. Here, we obtain $\mathbf{Po}_1^L = (x_1^L, y_1^L, z_1^L)$ and $\mathbf{Po}_1^R = (x_1^R, y_1^R, z_1^R)$. Thus, we can obtain the following:

$$\begin{aligned} \mathbf{C}_n^{b_i} &= \begin{bmatrix} \cos \gamma_i & 0 & -\sin \gamma_i \\ 0 & 1 & 0 \\ \sin \gamma_i & 0 & \cos \gamma_i \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_i & \sin \theta_i \\ 0 & -\sin \theta_i & \cos \theta_i \end{bmatrix} \begin{bmatrix} \cos \varphi_i & -\sin \varphi_i & 0 \\ \sin \varphi_i & \cos \varphi_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos \gamma_i \cos \varphi_i + \sin \gamma_i \sin \varphi_i \sin \theta_i & -\cos \gamma_i \sin \varphi_i + \sin \gamma_i \cos \varphi_i \sin \theta_i & -\sin \gamma_i \cos \theta_i \\ \sin \varphi_i \cos \theta_i & \cos \varphi_i \cos \theta_i & \sin \theta_i \\ \sin \gamma_i \cos \varphi_i - \cos \gamma_i \sin \varphi_i \sin \theta_i & -\sin \gamma_i \sin \varphi_i - \cos \gamma_i \cos \varphi_i \sin \theta_i & \cos \gamma_i \cos \theta_i \end{bmatrix}, \end{aligned} \quad (1)$$

where $\mathbf{C}_n^{b_j}, j \in [0, 8]$ indicates the rotation matrix from the b-frame of the j th point to the n-frame. θ, γ , and φ are pitch, roll, and yaw, respectively. Then, we can compute the joint's position of the left and right upper arm (denoted as $\mathbf{Po}_2^L$ and $\mathbf{Po}_2^R$) using the following equations:

$$\mathbf{C}_{b_1}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_1}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_1})^T, \quad (2)$$

$$\mathbf{Po}_2^L = l_E^L \mathbf{C}_{b_1,2}^{b_0} + \mathbf{Po}_1^L, \quad (3)$$

$$\mathbf{C}_{b_1^R}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_2^R}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_2^R})^T, \quad (4)$$

$$\mathbf{Po}_2^R = l_E^R \mathbf{C}_{b_1^R,2}^{b_0} + \mathbf{Po}_1^R. \quad (5)$$

Similarly, the position of the left and right wrist joints (denoted as $\mathbf{Po}_3^L$ and $\mathbf{Po}_3^R$) can be computed using the following equations:

$$\mathbf{C}_{b_3^L}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_3^L}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_3^L})^T, \quad (6)$$

$$\mathbf{Po}_3^L = l_W^L \mathbf{C}_{b_3^L,2}^{b_0} + \mathbf{Po}_2^L, \quad (7)$$

$$\mathbf{C}_{b_3^R}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_3^R}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_3^R})^T, \quad (8)$$

$$\mathbf{Po}_3^R = l_W^R \mathbf{C}_{b_3^R,2}^{b_0} + \mathbf{Po}_2^R. \quad (9)$$

We can compute the joint's position of the left and right knee (denoted as $\mathbf{Po}_4^L$ and $\mathbf{Po}_4^R$) using the following equations:

$$\mathbf{C}_{b_4^L}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_4^L}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_4^L})^T, \quad (10)$$

$$\mathbf{Po}_4^L = l_k^L \mathbf{C}_{b_4^L,2}^{b_0} + \mathbf{Po}_6^L, \quad (11)$$

$$\mathbf{C}_{b_4^R}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_4^R}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_4^R})^T, \quad (12)$$

$$\mathbf{Po}_4^R = l_k^R \mathbf{C}_{b_4^R,2}^{b_0} + \mathbf{Po}_6^R. \quad (13)$$

Then, we can compute the joint's position of the left and right ankle (denoted as $\mathbf{Po}_5^L$ and $\mathbf{Po}_5^R$) using the following equations:

$$\mathbf{C}_{b_5^L}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_5^L}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_5^L})^T, \quad (14)$$

$$\mathbf{Po}_5^L = l_a^L \mathbf{C}_{b_5^L,2}^{b_0} + \mathbf{Po}_4^L, \quad (15)$$

$$\mathbf{C}_{b_5^R}^{b_0} = \mathbf{C}_n^{b_0} \mathbf{C}_{b_5^R}^n = \mathbf{C}_n^{b_0} (\mathbf{C}_n^{b_5^R})^T, \quad (16)$$

$$\mathbf{Po}_5^R = l_a^R \mathbf{C}_{b_5^R,2}^{b_0} + \mathbf{Po}_4^R. \quad (17)$$

where $\mathbf{C}_{b_j^L,2}^{b_0}, \mathbf{C}_{b_j^R,2}^{b_0}, j \in [0, 8]$ represents the first column of $\mathbf{C}_{b_j^L}^{b_0}, \mathbf{C}_{b_j^R}^{b_0}, j \in [0, 8]$.

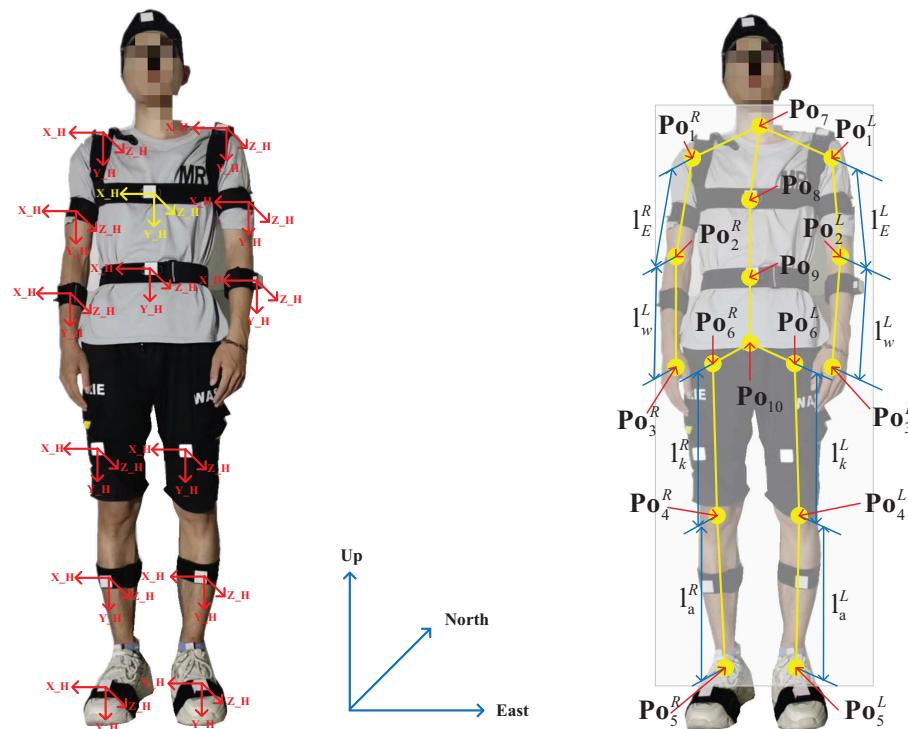


Figure 2. Coordinate system and key parameters of the human body used in this work.

3. ELM/FIR Integrated Filtering

In this section, we design ELM/FIR filtering, as shown in Figures 1 and 2. First, we introduce the scheme of the ELM/FIR filter. Second, a data fusion model is derived. Third, the ELM and FIR methods designed based on the data fusion model are presented.

3.1. Scheme of the ELM/FIR-Integrated Filtering

In this work, we divide the combined filtering algorithm into two stages—training and prediction stages—when the Kinect data are available and unavailable, respectively.

The scheme of the ELM/FIR integrated filtering when Kinect data are available is shown in Figure 3. In this stage, we employ three FIR filters and two ELM methods. First, both the IMU and Kinect measure the joint position in parallel. Then, FIR filters 1 and 2 are used to estimate the positions of the IMU and Kinect. Thereafter, their estimations are fused using the IMM and Rauch–Tung–Striebel (R–T–S) smoothing method. Then, the output of the RTS smoothing method is used to compute the positions. Meanwhile, IMU’s and Kinect’s solutions are, respectively, used as the input and target of ELM 1, which is used to build the mapping between the IMU-measured position and the Kinect-measured position. The output of ELM 1 is used as the measurement of FIR filtering 3. ELM 2 employs the outputs of ELM 1 as input and the difference between the RTS method’s output and FIR filtering 3’s output as the target, which is used to build the mapping between them.

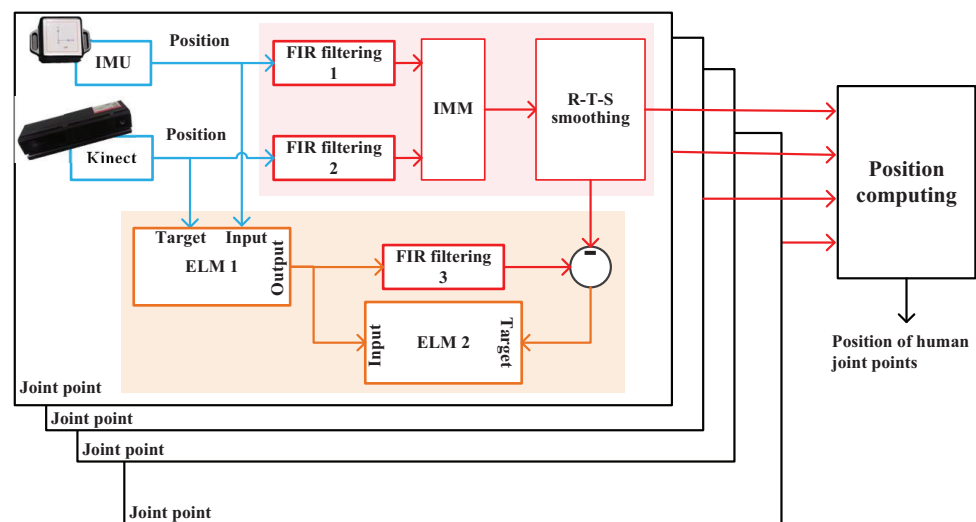


Figure 3. Scheme of the ELM/FIR-integrated filtering when Kinect data are available.

The scheme of the ELM/FIR-integrated filtering when Kinect data are unavailable is shown in Figure 4. In this stage, because we cannot obtain Kinect data, the IMU-measured position is input to ELM 1 directly. Then, ELM 1 outputs its estimated position with the mapping built in the training stage, which is employed by FIR filtering 3 as its measurement. Then, ELM 2 works to estimate the corresponding error, which is used to correct FIR filtering 3’s solution.

With all the joint points’ estimations, the positions can be computed. Notably, the Kinect data of each node may not necessarily be lost simultaneously.

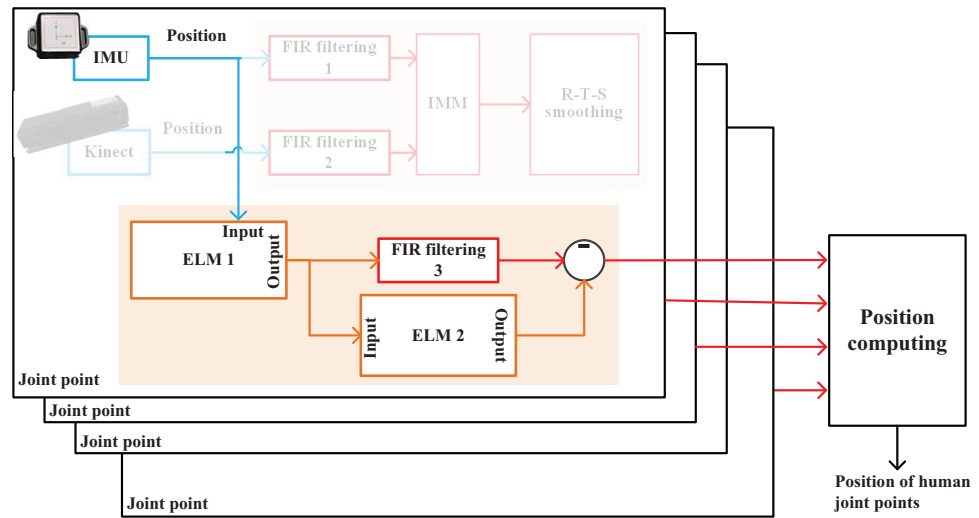


Figure 4. Scheme of the ELM/FIR-integrated filtering when Kinect data are unavailable.

3.2. Data Fusion Model of FIR Filtering

From the scheme mentioned above, we can see that three FIR filters are used in this work. The state equation used in this work can be given as follows:

$$\underbrace{\begin{bmatrix} Lx_t^{m,j-} \\ vx_t^{m,j-} \\ Ly_t^{m,j-} \\ vy_t^{m,j-} \end{bmatrix}}_{\mathbf{L}_t^{m,j-}} = \underbrace{\begin{bmatrix} 1 & \delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{T}^{m,j}} \underbrace{\begin{bmatrix} Lx_t^{m,j} \\ vx_t^{m,j} \\ Ly_t^{m,j} \\ vy_t^{m,j} \end{bmatrix}}_{\mathbf{L}_t^{m,j}} + \mathbf{w}_t^{m,j}, \quad (18)$$

where $(Lx_t^{m,j-}, Ly_t^{m,j-})$ is the j th joint's position at the time index t , $(vx_t^{m,j-}, vy_t^{m,j-})$ is the j th joint's velocity at the time index t , m denotes the m th FIR filter used for the j th joint, δt is the sampling time, and $\mathbf{w}_t^{m,j} \sim \mathcal{N}(0, \mathbf{Q})$ is the system noise.

$$\underbrace{\begin{bmatrix} L\hat{x}_t^{m,j} \\ L\hat{y}_t^{m,j} \end{bmatrix}}_{\mathbf{Z}_t^{m,j}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{G}} \mathbf{L}_t^{m,j-} + \mathbf{v}_t^{m,j}, \quad (19)$$

where $(L\hat{x}_t^{m,j}, L\hat{y}_t^{m,j})$, $m \in [1, 3]$ is the j th joint's position measured using IMU, Kinect, and ELM 1's output and $\mathbf{v}_t^{m,j} \sim \mathcal{N}(0, \mathbf{R})$ is the measurement noise.

3.3. IMM-FIR Filtering

Based on models (18) and (19), the FIR filtering method used in this paper can be listed as Algorithm 1. In Algorithm 1, $D^{m,j}$ is the dimension of $\mathbf{L}_t^{m,j}$, and $L^{m,j}$ is the filtering size. To the m th FIR filter, this algorithm has a dead zone. In this work, when the time index is in the dead zone, we employ the Kalman filter to replace the FIR's work (lines 3–9 in Algorithm 1). When the time index is greater than the $L_t^{m,j}$, the FIR filter runs the one-step prediction by using the following equations:

$$\mathbf{L}_{jj}^{m,j-} = \mathbf{T}^{m,j} \mathbf{L}_{jj}^{m,j} + \mathbf{w}_{jj}^{m,j}, \quad (20)$$

Then, the gain can be computed by using the following equation:

$$\mathbf{H}_{jj}^{m,j} = \left(\mathbf{G}\mathbf{G}^T + \left(\mathbf{T}^{m,j} \mathbf{H}_{jj-1}^{m,j} \mathbf{T}^{m,j} \right)^{-1} \right)^{-1}, \quad (21)$$

$$\mathbf{K}_{jj}^{m,j} = \mathbf{H}_{jj}^{m,j} \mathbf{G}^T. \quad (22)$$

Then, the measurement update can be computed:

$$\mathbf{L}_{jj}^{m,j} = \mathbf{L}_{jj}^{m,j-} + \mathbf{K}_{jj}^{m,j} \left(\mathbf{Z}_{jj}^{m,j} - \mathbf{G}\mathbf{L}_{jj}^{m,j-} \right). \quad (23)$$

Note that the subscript jj is the counting of internal loops in FIR filter. The FIR filter employs the recent measurement to estimate the robust output.

Algorithm 1: m th FIR filter used for the j th joint's position in this work

Data: $\mathbf{Z}_t^{m,j}$, $\mathbf{L}_0^{m,j}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{P}_0^{m,j}$, $D^{m,j}$, $L^{m,j}$
Result: $\hat{\mathbf{s}}_l$, $\hat{\mathbf{O}}_l$

```

1 begin
2   for  $t = 1 : \infty$  do
3     if  $t < L^{m,j}$  then
4        $\mathbf{L}_t^{m,j-} = \mathbf{T}^{m,j} \mathbf{L}_{t-1}^{m,j}$ ;
5        $\mathbf{P}_t^{m,j-} = \mathbf{T}^{m,j} \mathbf{P}_{t-1}^{m,j} (\mathbf{T}^{m,j})^T + \mathbf{Q}$ ;
6        $\mathbf{K}_t^{m,j} = \mathbf{P}_t^{m,j-} \mathbf{G}^T \left( \mathbf{G} \mathbf{P}_t^{m,j-} \mathbf{G}^T + \mathbf{R} \right)^{-1}$ ;
7        $\mathbf{L}_t^{m,j} = \mathbf{L}_t^{m,j-} + \mathbf{K}_t^{m,j} \left( \mathbf{Z}_t^{m,j} - \mathbf{G} \mathbf{L}_t^{m,j-} \right)$ ;
8        $\mathbf{P}_t^{m,j} = \left( \mathbf{I} - \mathbf{K}_t^{m,j} \mathbf{G} \right) \mathbf{P}_t^{m,j-}$ ;
9     else
10       $m = t - L^{m,j} + 1$ ,  $jj = m + D^{m,j} - 1$ ;
11       $\tilde{\mathbf{L}}_{jj} = \hat{\mathbf{L}}_{jj}^{m,j}$ ;
12       $\tilde{\mathbf{P}}_{jj} = \hat{\mathbf{P}}_{jj}^{m,j}$ ;
13      for  $jj = m + D^{m,j} : t$  do
14         $\mathbf{L}_{jj}^{m,j-} = \mathbf{T}^{m,j} \mathbf{L}_{jj}^{m,j} + \mathbf{w}_{jj}^{m,j}$ ;
15         $\mathbf{H}_{jj}^{m,j} = \left( \mathbf{G}\mathbf{G}^T + \left( \mathbf{T}^{m,j} \mathbf{H}_{jj-1}^{m,j} \mathbf{T}^{m,j} \right)^{-1} \right)^{-1}$ ;
16         $\mathbf{K}_{jj}^{m,j} = \mathbf{H}_{jj}^{m,j} \mathbf{G}^T$ ;
17         $\mathbf{L}_{jj}^{m,j} = \mathbf{L}_{jj}^{m,j-} + \mathbf{K}_{jj}^{m,j} \left( \mathbf{Z}_{jj}^{m,j} - \mathbf{G}\mathbf{L}_{jj}^{m,j-} \right)$ ;
18      end for
19       $\hat{\mathbf{L}}_t^{m,j} = \tilde{\mathbf{L}}_{jj}^{m,j}$ ;
20       $\hat{\mathbf{P}}_t^{m,j} = \tilde{\mathbf{P}}_{jj}^{m,j}$ ;
21    end if
22  end for
23 end
```

Figure 3 shows that we employ FIR filters 1 and 2 to measure the j th joint's position $\mathbf{L}_t^{1,j}$ and $\mathbf{L}_t^{2,j}$ using IMU and Kinect's data, respectively. Then, we introduce the *Markov* matrix, which can be given as follows:

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}. \quad (24)$$

Thus, we can determine that the predicted probability (normalization constant) $\bar{c}^m$ of model i to model r is:

$$\bar{c}^m = \sum_{i=1}^r \Gamma_i^m \mu_i^{m,j-}, m \in [1, 2]. \quad (25)$$

The mixed probability $\mu_i^{m,j-}$ from model i to model r is:

$$\mu_i^{m,j-} = \sum_{i=1}^r \Gamma_i \mu_i^{m,j-} / \bar{c}^m, m \in [1, 2], \quad (26)$$

where μ is the model probability for each filter and Γ is the probability transfer matrix.

With the FIR filters 1 and 2, we can determine that the blended state estimation of the model $\hat{\mathbf{L}}_0^{m,j-}$ and the mixed covariance estimation of the model $\mathbf{P}_0^{m,j-}$:

$$\hat{\mathbf{L}}_0^{m,j-} = \sum_{i=1}^r \hat{\mathbf{L}}_i^{m,j-} \mu_i^{m,j-}, \quad (27)$$

$$\mathbf{P}_0^{m,j-} = \sum_{i=1}^r \mu_i^{m,j-} (\mathbf{P}_i^{m,j-} + (\hat{\mathbf{L}}_i^{m,j-} - \hat{\mathbf{L}}_0^{m,j-})(\hat{\mathbf{L}}_i^{m,j-} - \hat{\mathbf{L}}_0^{m,j-})^T). \quad (28)$$

Take $\hat{\mathbf{L}}^{m,j}$, $\mathbf{P}^{m,j-}$, and $\mathbf{Z}_t$ as inputs to update prediction state $\hat{\mathbf{L}}_t^{m,j}$ and filter covariance $\mathbf{P}_t^{m,j}$:

$$\hat{\mathbf{L}}^{1,j} = \mathbf{T}^{1,j-} \hat{\mathbf{L}}_0^{1,j-}, \quad (29)$$

$$\hat{\mathbf{L}}^{2,j} = \mathbf{T}^{2,j-} \hat{\mathbf{L}}_0^{2,j-}, \quad (30)$$

$$\mathbf{P}^{1,j-} = \mathbf{T}^{1,j-} \mathbf{P}_0^{1,j-} [\mathbf{T}^{1,j-}]^T + \mathbf{Q}, \quad (31)$$

$$\mathbf{P}^{2,j-} = \mathbf{T}^{2,j-} \mathbf{P}_0^{2,j-} [\mathbf{T}^{2,j-}]^T + \mathbf{Q}, \quad (32)$$

$$\mathbf{K}_t^1 = \mathbf{P}^{1,j-} \mathbf{G}^T (\mathbf{G} \mathbf{P}^{1,j-} \mathbf{G}^T + \mathbf{R})^{-1}, \quad (33)$$

$$\mathbf{K}_t^2 = \mathbf{P}^{2,j-} \mathbf{G}^T (\mathbf{G} \mathbf{P}^{2,j-} \mathbf{G}^T + \mathbf{R})^{-1}, \quad (34)$$

$$\hat{\mathbf{L}}_t^{1,j} = \hat{\mathbf{L}}^{1,j-} + \mathbf{K}_t^1 (\mathbf{Z}_t^{1,j} - \mathbf{G} \hat{\mathbf{L}}^{1,j-}), \quad (35)$$

$$\hat{\mathbf{L}}_t^{2,j} = \hat{\mathbf{L}}^{2,j-} + \mathbf{K}_t^2 (\mathbf{Z}_t^{2,j} - \mathbf{G} \hat{\mathbf{L}}^{2,j-}), \quad (36)$$

$$\mathbf{P}_t^{1,j} = (\mathbf{I} - \mathbf{K}_t^1 \mathbf{G}) \mathbf{P}^{1,j-}, \quad (37)$$

$$\mathbf{P}_t^{2,j} = (\mathbf{I} - \mathbf{K}_t^2 \mathbf{G}) \mathbf{P}^{2,j-}. \quad (38)$$

The likelihood function Λ_t^i is used to update the model probability μ_t , calculated as follows:

$$\Lambda_t^1 = \frac{1}{(2\pi)^{n/2} |\mathbf{S}_t^1|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{v}_t^1)^T (\mathbf{S}_t^1)^{-1} \mathbf{v}_t^1 \right\}, \quad (39)$$

$$\Lambda_t^2 = \frac{1}{(2\pi)^{n/2} |\mathbf{S}_t^2|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{v}_t^2)^T (\mathbf{S}_t^2)^{-1} \mathbf{v}_t^2 \right\}, \quad (40)$$

where $\mathbf{v}_t^i$ and $\mathbf{S}_t^i$ are the measurement error and measurement error covariance matrix, calculated as follows:

$$\mathbf{v}_t^1 = \mathbf{Z}_t^{1,j} - \mathbf{T}^{1,j} \hat{\mathbf{L}}_{t,j}^1, \quad (41)$$

$$\mathbf{S}_t^1 = \mathbf{G} \mathbf{P}^{1,j} \mathbf{G}^T + \mathbf{R}, \quad (42)$$

$$\mathbf{v}_t^2 = \mathbf{Z}_t^{2,j} - \mathbf{T}^{1,j} \hat{\mathbf{L}}_{t,j}^{2,j}, \quad (43)$$

$$\mathbf{S}_t^2 = \mathbf{G} \mathbf{P}^{2,j} \mathbf{G}^T + \mathbf{R}, \quad (44)$$

where μ_t is the credibility after fusion and c is the denominator coefficient to achieve normalization, calculated as follows:

$$c = \sum_{m=1}^r \Lambda_t^m \bar{c}^m \quad (45)$$

$$\mu_t^1 = \Lambda_t^1 \bar{c} / c. \quad (46)$$

Based on the updated confidence level μ_t^m , the models can be fused sequentially to obtain the output target state and covariance matrix of the algorithm:

$$\hat{\mathbf{L}}_t^{m,j} = \sum_{m=1}^r \hat{\mathbf{L}}_{t,j}^{m,j} \mu_t^m, \quad (47)$$

$$\mathbf{P}_t^{m,j} = \sum_{j=1}^r \mu_t^m \left(\mathbf{P}_j^{m,j} + (\hat{\mathbf{L}}_j^{m,j} - \hat{\mathbf{L}}_{t,j}^{m,j}) (\hat{\mathbf{L}}_j^{m,j} - \hat{\mathbf{L}}_{t,j}^{m,j})^T \right). \quad (48)$$

3.4. ELM Method

The ELM method used in this work is similar to the method we proposed previously [42]. In this section, we introduce this method briefly. On the basis of the data fusion model mentioned above, we can compute ELM's activation function $\chi(\cdot)$ as follows:

$$\sum_{q=1}^d \alpha_q \chi(\beta_q \mathbf{I}_e + b_q) = \delta_e, e \in [1, s-1], \quad (49)$$

where $\mathbf{I}_e = [I_1, I_2, \dots, I_{s-1}]$. Thus, we can obtain the following equation:

$$\sum_{q=1}^d \alpha_q \chi(\beta_q \mathbf{I}_e + b_q) = y_e, e \in [1, s-1]. \quad (50)$$

Then, (50) can be rewritten as

$$\underbrace{\begin{bmatrix} f_1^E(I_1) \\ \vdots \\ f_{s-1}^E(I_{s-1}) \end{bmatrix}}_{\mathbf{F}^E} \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{s-1} \end{bmatrix}}_{\boldsymbol{\beta}^E} = \underbrace{\begin{bmatrix} z_1 \\ \vdots \\ z_{s-1} \end{bmatrix}}_{\mathbf{z}^E}, \quad (51)$$

and $\boldsymbol{\beta}^E$ can be computed using the following equation:

$$\hat{\boldsymbol{\beta}}^E = \mathbf{F}^{E+} \mathbf{z}^E. \quad (52)$$

4. Test

In this work, we perform one real test to show the effectiveness of the proposed method. First, the corresponding experimental parameters are presented. Then, the performance of the proposed method is investigated.

4.1. Experimental Parameters

Experimental parameters are introduced in this section. Figure 5 shows the block diagram of the experimental system. Herein, we employed eight IMUs fixed on the target human to measure the corresponding posture. Meanwhile, Kinect was also used to measure the target human's joint positions. All the sensors' data were collected using a computer. In this work, the model of the IMU is ICM-20948, and the IMU parameters are given in Table 1. Only eight joint positions were considered herein, as shown in Figure 1. Meanwhile, we employed Kinect 2.0 as the visual sensor in this work; its parameters are given in Table 2. The test environment is shown in Figure 6. When we performed the test, the target human moved in front of Kinect and the IMUs were fixed on the target human's body. The human moved according to predetermined movements, and the computer collected all the sensor data.

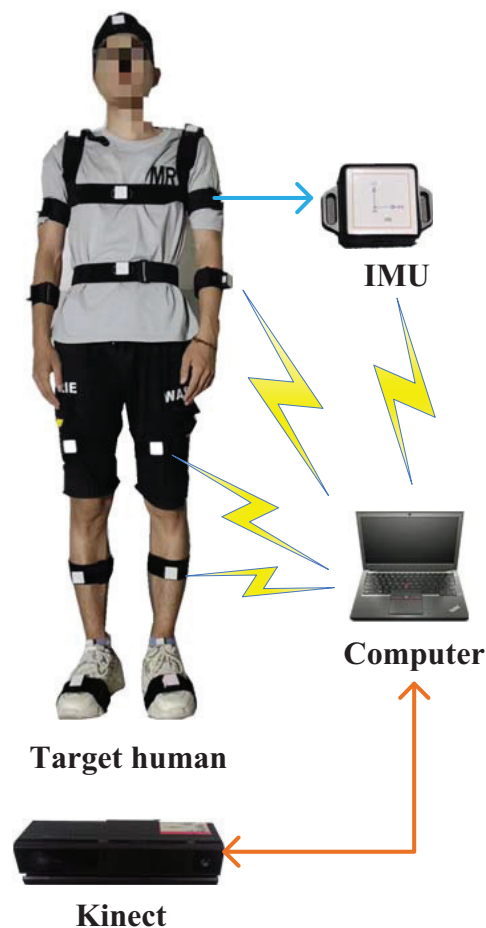


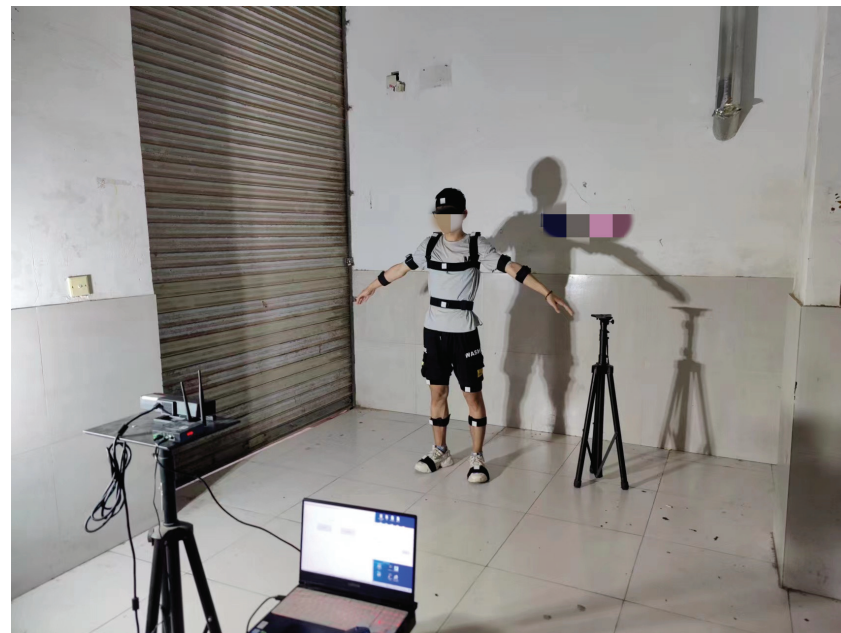
Figure 5. Block diagram of the experimental system.

Table 1. Parameters of the IMUs used in the test.

Parameter	Value
Sensor's precision	0.01° (pitch and roll), 0.1° (yaw)
Sampling frequency	100 Hz
Measurement dimension	3
Data transmission distance	100 m
Working voltage	4.2 V

Table 2. Parameters of Kinect used in the test.

Parameter	Value
Resolution of color image frames	1920 × 1080
Resolution of deep frames	512 × 424
Detectable range	0.5–4.5 m
Resolution of infrared image frames	512 × 484
Field of view	70° × 60°

**Figure 6.** Test environment.

4.2. Positioning Accuracy

In this section, six joints' positioning accuracies measured using the proposed method were investigated. Herein, we employed the solutions of IMU, Kinect, and ELM/KF filtering to compare their performances. Figure 7 displays the positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the right-arm elbow. In this figure, the black line represents the solution of the ELM/FIR method, the blue line represents the solution of the ELM/KF method, the green line represents the solution of the IMU-only solution, the pink line represents the solution of the Kinect-only solution, and the red line represents the reference value. The figure shows that in this joint, the performances of ELM/FIR and ELM/KF were similar, and the positions estimated using the methods mentioned above were between the outputs of the Kinect and IMU. It can be seen that in the figure, the blue and black lines are almost identical in all three directions. In the figure, it can be seen that there are some protrusions in the solution values of the ELM/FIR and ELM/KF filters, which indicate that the Kinect data are not available in these time indexes. However, when Kinect data were unavailable, the output of the proposed ELM/FIR method was closer to

the solution of the ELM/KF method, which deduced that the proposed ELM/FIR method was effective in maintaining the performance of the filter during Kinect data outage.

The cumulative distribution functions (CDFs) of the position errors measured using ELM/FIR and ELM/KF in the right-arm elbow are shown in Figure 8. In this figure, the red line represents the solution of the ELM/FIR method, the blue line represents the solution of the ELM/KF method. As shown in this figure, the proposed ELM/FIR method exhibited better performance than that of ELM/KF. At 0.9, the proposed ELM/FIR filter could reduce the localization error from 0.0866 to 0.0678 m, improving the localization accuracy by approximately 12.71%, which indicates that the proposed ELM/FIR is more effective than the proposed ELM/KF method in this joint.

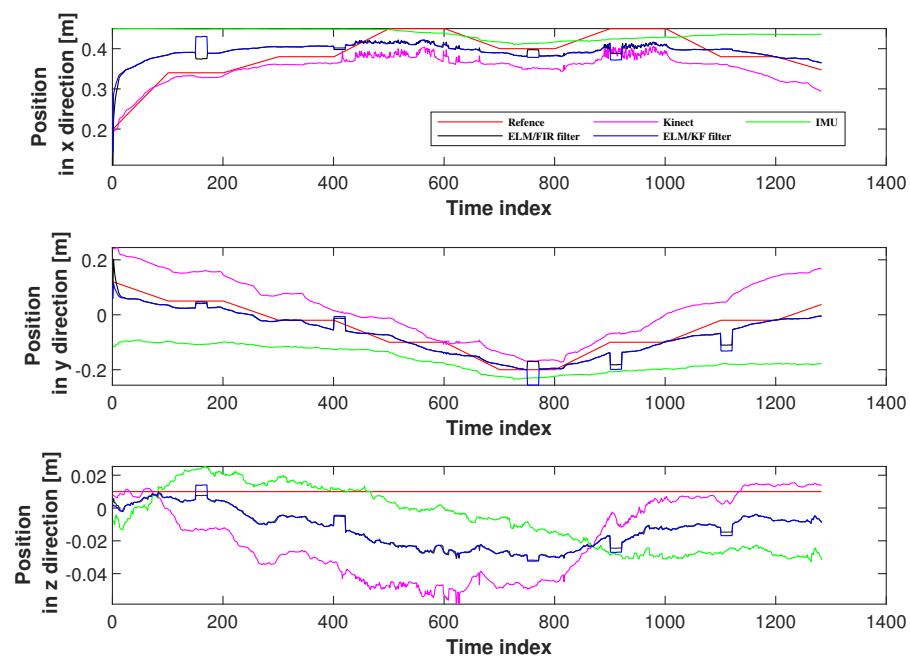


Figure 7. Positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the right-arm elbow.

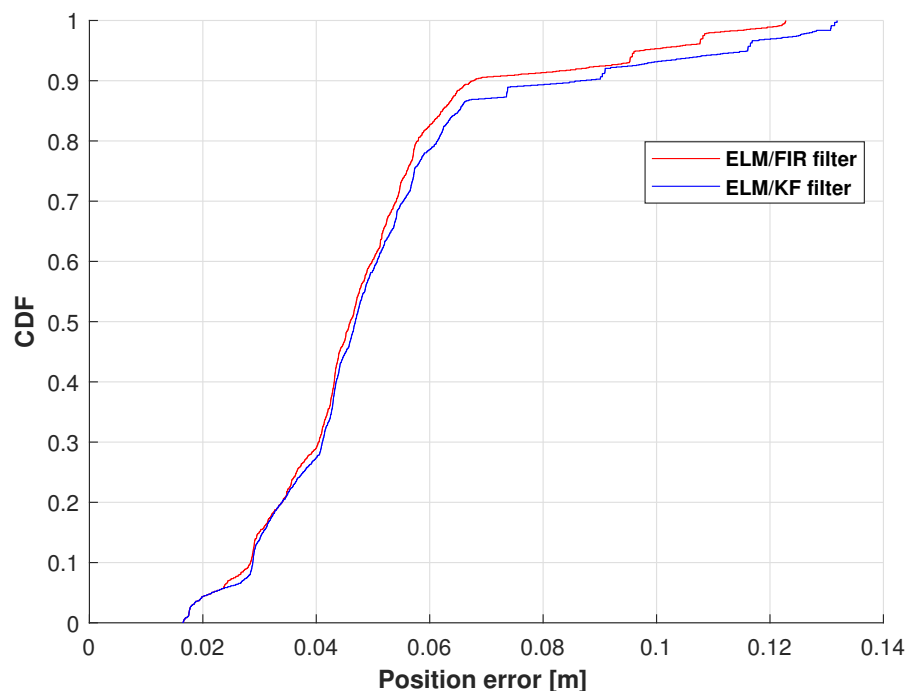


Figure 8. CDF of the position errors measured using ELM/FIR and ELM/KF in the right-arm elbow.

The positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the right-arm wrist are listed in Figure 9. Similar to the case of the right-arm elbow, when the data of IMU and Kinect are available, the performance of the ELM/FIR and ELM/KF are almost the same. And when the Kinect data experience an outage, it can be seen that the black lines are closer to the reference value compared to the blue line, which means that the proposed method is more effective in reducing the localization error when Kinect data are unavailable. Figure 10 displays the CDF of the position errors measured using ELM/FIR and ELM/KF in the right-arm wrist. The proposed method thus improved the localization error by approximately 16.40%. Moreover, the root-mean-squared errors (RMSEs) measured using the ELM/FIR and ELM/KF filters in the right-arm wrist are listed in Table 3. The table shows that the ELM/FIR method improved the localization error from 0.0690 to 0.0641 m compared to the ELM/KF method. From the figures and the table mentioned above, it can be seen that the proposed ELM/FIR is more effective than the proposed ELM/KF method in this joint.

The CDFs of the position errors measured using ELM/FIR and ELM/KF in the left-arm elbow are shown in Figure 11. Therein, the position error of ELM/FIR is smaller than that of ELM/KF. This shows that the proposed method is effective in reducing the positioning error.

Table 3. RMSEs measured using the ELM/FIR and ELM/KF filters in the right-arm wrist.

Methods	X (m)	Y (m)	Z (m)	Mean (m)
ELM/KF filter	0.0959	0.0819	0.0292	0.0690
ELM/FIR filter	0.0903	0.0739	0.0282	0.0641

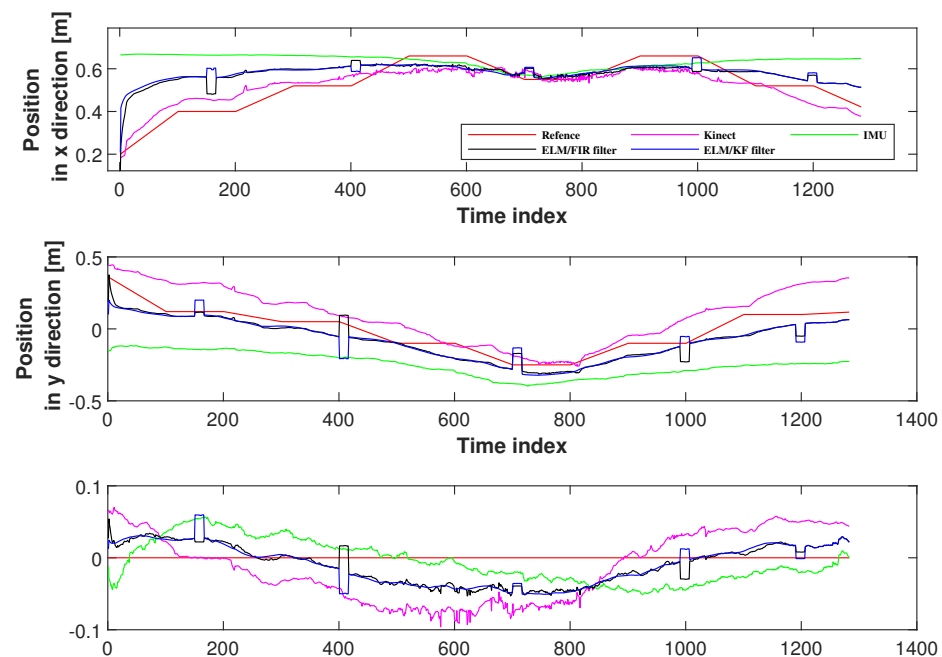


Figure 9. Positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the right-arm wrist.

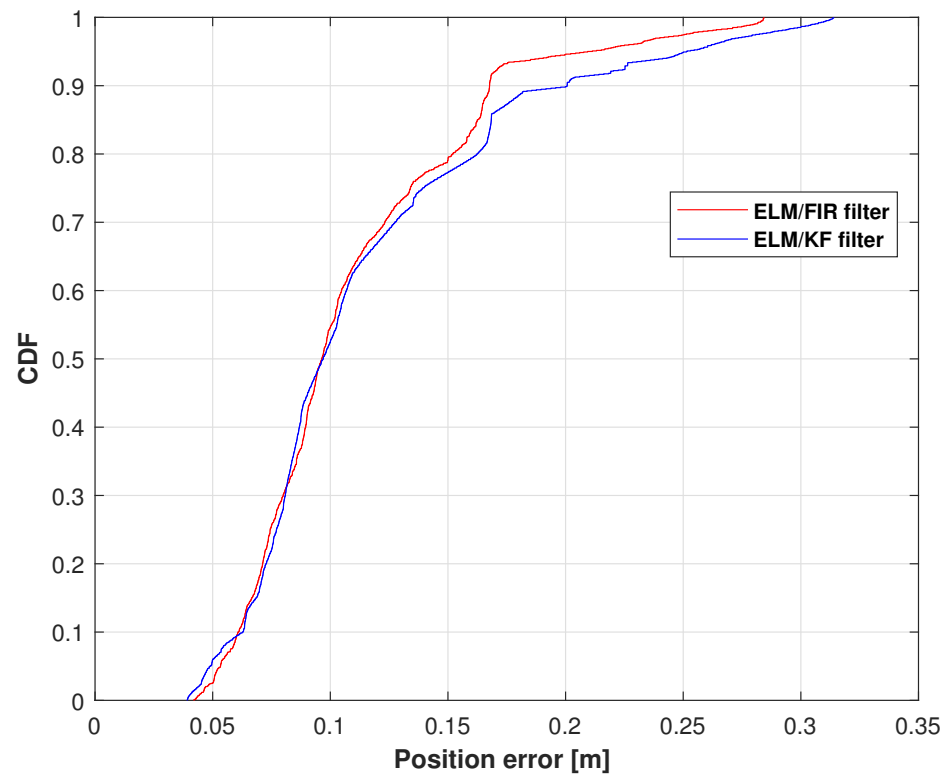


Figure 10. CDF of the position errors measured using ELM/FIR and ELM/KF in the right-arm wrist.

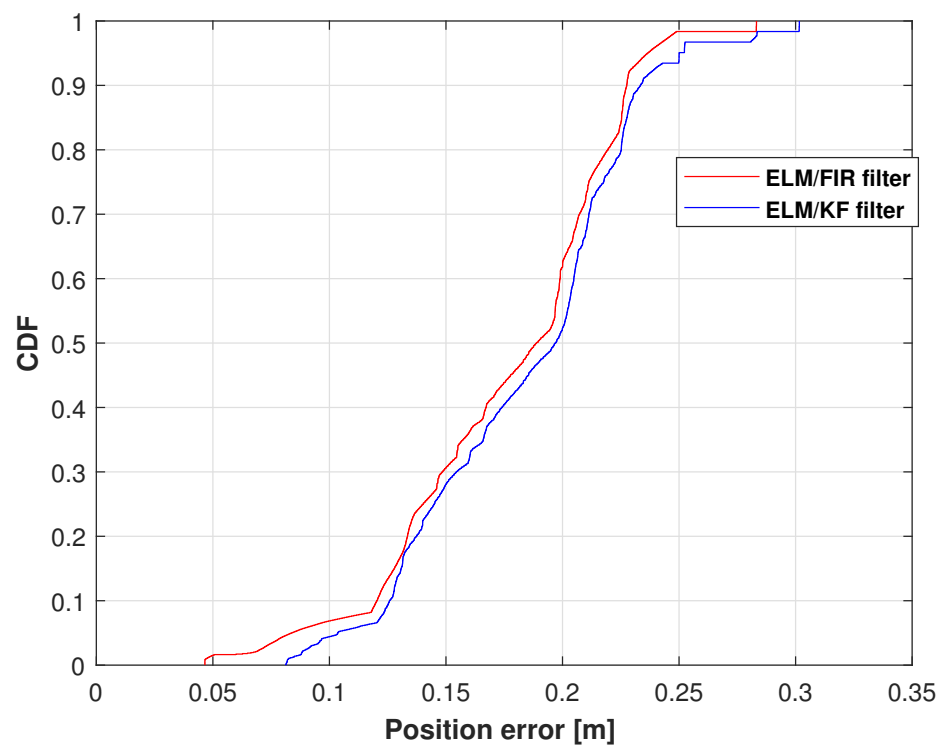


Figure 11. CDF measured using ELM + FIR and ELM + KF in the left-arm elbow.

The positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the left-arm wrist are given in Figure 12. Similar to the case of the right-arm elbow, the proposed method is more effective in reducing the localization error when Kinect data are unavailable. Figure 13 displays the CDF of the position errors measured using ELM/FIR and ELM/KF in the left-arm wrist. The proposed method thus improved the localization error by

approximately 56.71%. Moreover, the RMSEs measured using the ELM/FIR filter and ELM/KF filter in the left-arm wrist are presented in Table 4. According to this table, the ELM/FIR method improved the localization error from 0.1493 to 0.0607 m compared with the ELM/KF method.

Table 4. RMSEs measured using the ELM/FIR filter and ELM/KF filter in the left-arm wrist.

Methods	X (m)	Y (m)	Z (m)	Mean (m)
ELM/KF filter	0.2023	0.2103	0.0352	0.1493
ELM/FIR filter	0.0617	0.0873	0.0333	0.0607

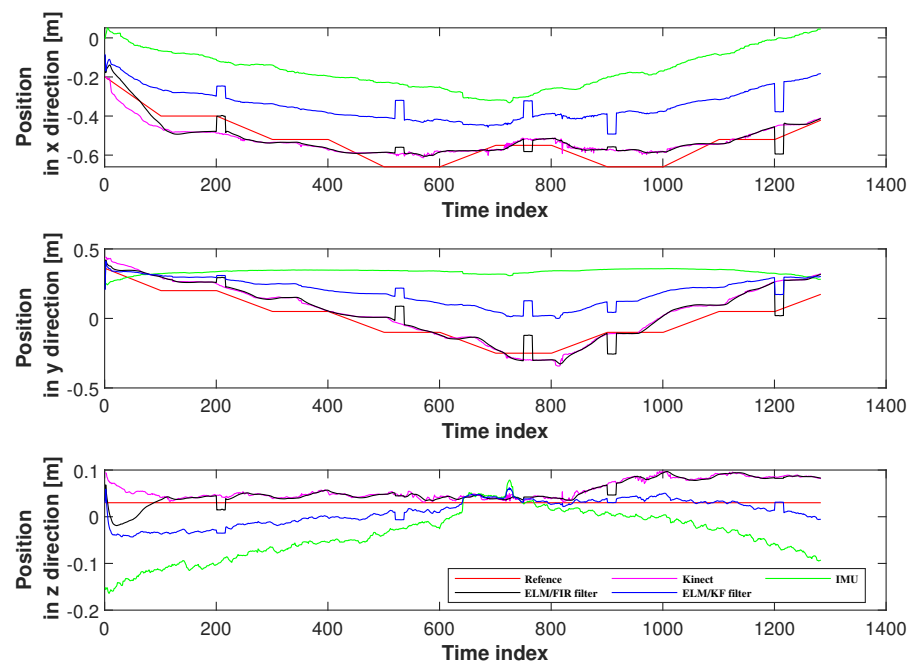


Figure 12. Positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the left-arm wrist.

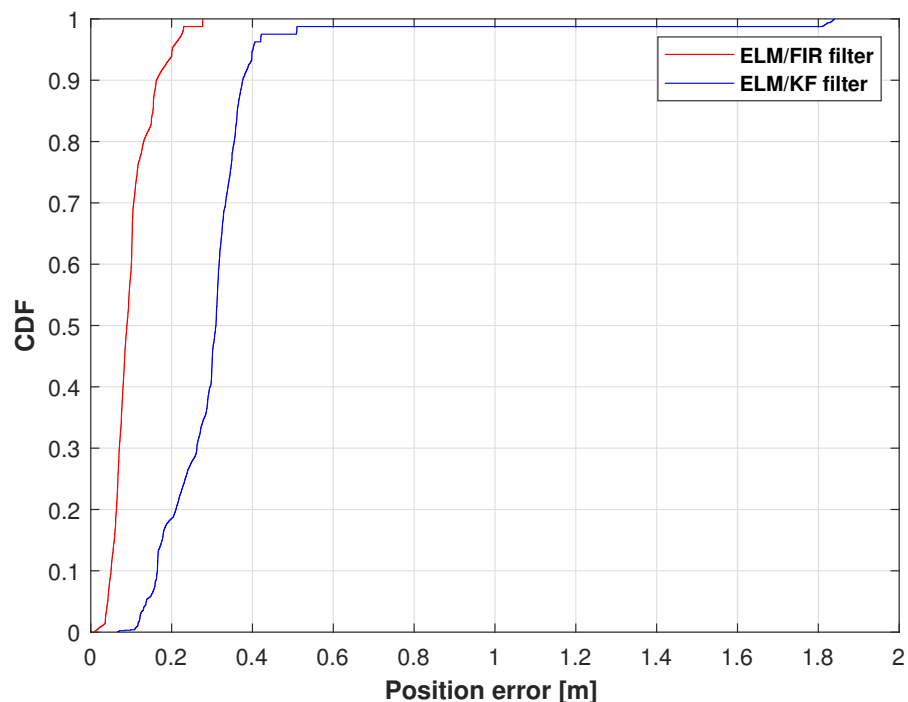


Figure 13. CDF measured using ELM + FIR and ELM + KF in the left-arm wrist.

The positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the right knee are shown in Figure 14. Similar to the case of the right-arm elbow, the proposed method is more effective in reducing the localization error when Kinect data are unavailable. Figure 15 displays the CDF of the position errors measured using ELM/FIR and ELM/KF in the right knee. The proposed method improved the localization error by approximately 48.32%. Moreover, the RMSEs measured using the ELM/FIR and ELM/KF filters in the right knee are given in Table 5. The table shows that the ELM/FIR method improved the localization error from 0.1128 to 0.0566 m compared with the ELM/KF method.

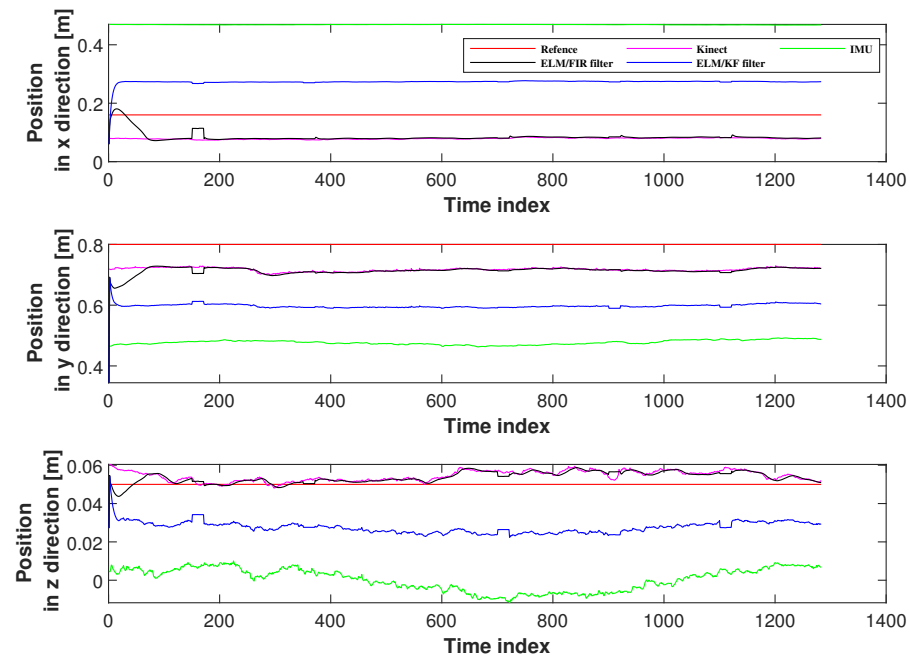


Figure 14. Positions measured using Kinect, IMU, ELM/FIR, and ELM/KF in the right knee.

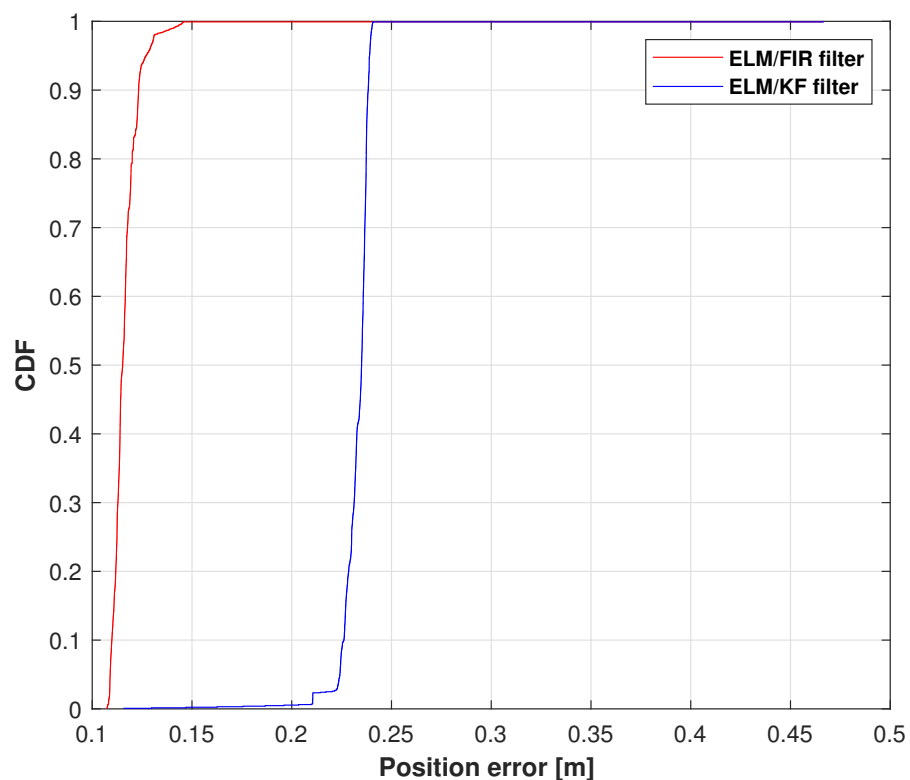
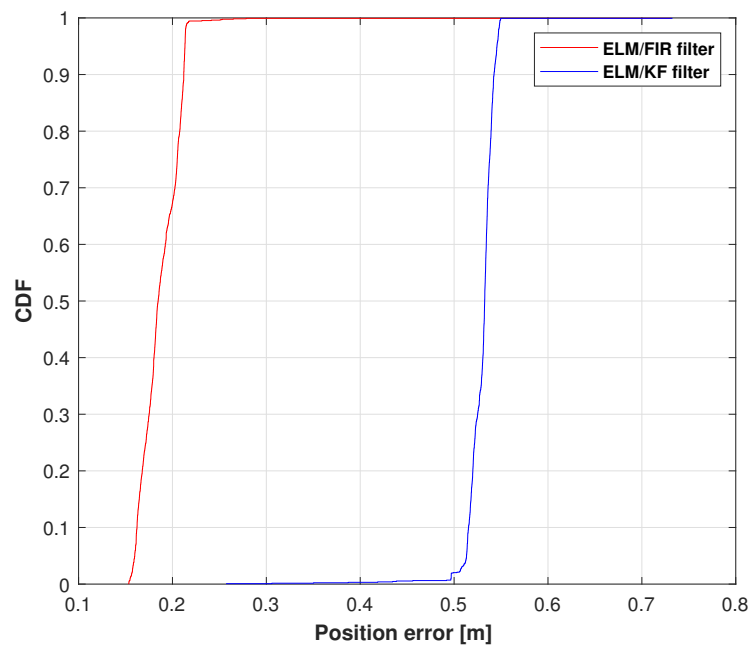


Figure 15. CDF measured using the ELM/FIR and ELM/KF filters in the right knee.

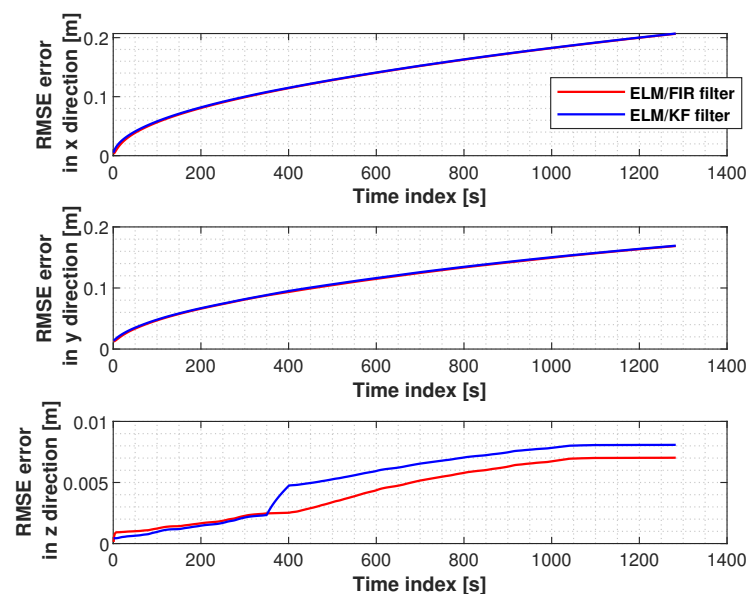
Table 5. RMSEs measured using the ELM/FIR and ELM/KF filters in the right knee.

Methods	X (m)	Y (m)	Z (m)	Mean (m)
ELM/KF filter	0.1133	0.2025	0.0226	0.1128
ELM/FIR filter	0.0771	0.0880	0.0047	0.0566

The position error CDFs measured using ELM/FIR and ELM/KF in the right ankle are presented in Figure 16. This figure shows that the proposed method's solution is closer to the reference value. Meanwhile, the CDF results show that the proposed method substantially improved the positioning accuracy.

**Figure 16.** CDF measured using the ELM/FIR and ELM/KF filters in the right ankle.

The RMSEs measured using the ELM/FIR and ELM/KF filters in the left knee are shown in Figure 17. Herein, the RMSEs of the two filters in the x and y directions are similar. However, the proposed method shows its effectiveness in the z direction.

**Figure 17.** RMSEs measured using the ELM/FIR and ELM/KF filters in the left knee.

From the analysis of the positioning accuracies of the different joints mentioned above, we can conclude that the proposed method is effective in reducing localization error, especially in Kinect outage areas.

5. Conclusions

To obtain accurate position information, a one-assistant method fusing the ELM/FIR filter and vision data was proposed herein for INS-based human motion capture. In the proposed method, when vision is available, the vision-based human position inputs to an FIR filter that accurately outputs the human position. Meanwhile, another FIR filter outputs the human position using INS data. Moreover, ELM was used to build mapping between the output of FIR and the corresponding error. When vision data were unavailable, FIR was used to provide the human posture and ELM was used to provide the estimation error built in the previously mentioned stage. In order to show the effectiveness of the proposed ELM/FIR filter, eight joint points are considered in this work. Test results show that the localization errors of the eight joint points measured using the proposed ELM/FIR filter are smaller than the values of the ELM/KF filter, especially when a Kinect data outage occurs, which demonstrates the effectiveness of the proposed method.

Author Contributions: Conceptualization, Y.X., A.Y. and T.S.; methodology, Y.X. and M.S.; software, R.G.; validation, R.G. and K.L.; formal analysis, Z.S.; investigation, Y.X.; resources, R.G.; data curation, Y.X.; writing—original draft preparation, Y.X.; writing—review and editing, Y.X.; visualization, Y.X.; supervision, Y.X.; project administration, Y.X.; funding acquisition, Y.X. and T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shandong Natural Science Foundation under Grants ZR2023MF121 and ZR2020MF067 and the 2022 Shandong Province Science and Technology Small and Medium Enterprises Innovation Ability Enhancement Project under Grant 2022TSGC2037.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Desmarais, Y.; Mottet, D.; Slangen, P.; Montesinos, P. A review of 3D human pose estimation algorithms for markerless motion capture. *Comput. Vis. Image Underst.* **2021**, *212*, 103275. [\[CrossRef\]](#)
- Wang, Z.; Gao, F.; Wu, Z.; Wang, D.; Guo, X.; Yu, S. A method for calculating lower extremity anatomical landmark trajectories based on inertial motion capture data. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 2734–2746. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wei, Y. Deep-learning-based motion capture technology in film and television animation production. *Secur. Commun. Netw.* **2022**, *2022*, 6040371. [\[CrossRef\]](#)
- Barak-Ventura, R.; Ruiz-Marín, M.; Nov, O.; Raghavan, P.; Porfiri, M. A low-cost telerehabilitation paradigm for bimanual training. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 395–406. [\[CrossRef\]](#)
- Skurowski, P.; Pawlyta, M. Detection and classification of artifact distortions in optical motion capture sequences. *Sensors* **2022**, *22*, 4076. [\[CrossRef\]](#)
- Lei, Y.; Deng, Y.; Dong, L.; Li, X.; Li, X.; Su, Z. A novel sensor fusion approach for precise hand tracking in virtual reality-based human—Computer interaction. *Biomimetics* **2023**, *8*, 326. [\[CrossRef\]](#)
- Han, Y. 2D-to-3D visual human motion converting system for home optical motion capture tool and 3-D smart TV. *IEEE Syst. J.* **2015**, *9*, 131–140.
- Chang, C.C.; Tsai, W.H. Vision-based tracking and interpretation of human leg movement for virtual reality applications. *IEEE Trans. Circuits Syst. Video Technol.* **2001**, *11*, 9–24. [\[CrossRef\]](#)
- Yu, T.; Zhao, J.; Li, Y.; Liu, Y. Towards robust and accurate single-view fast human motion capture. *IEEE Access* **2019**, *7*, 85548–85559. [\[CrossRef\]](#)
- Li, M.; Wei, F.; Li, Y.; Zhang, S.; Xu, G. Three-dimensional pose estimation of infants lying supine using data from a Kinect sensor with low training cost. *IEEE Sens. J.* **2021**, *21*, 6904–6913. [\[CrossRef\]](#)
- Zhao, G.; Zan, H.; Chen, J. Research on skeleton data compensation of gymnastics based on dynamic and static two-dimensional regression using Kinect. *Meas. Sci. Rev.* **2022**, *22*, 283–292. [\[CrossRef\]](#)
- Naeemabadi, M.; Dinesen, B.; Andersen, O.K.; Hansen, J. Influence of a marker-based motion capture system on the performance of microsoft Kinect v2 skeleton algorithm. *IEEE Sens. J.* **2019**, *19*, 171–179. [\[CrossRef\]](#)

13. Sohn, J.H.; Oh, S.; Lee, C.H.; Kim, S.S. Recursive inverse kinematics analysis for teaching human motion to a humanoid social robot using a depth camera. In Proceedings of the 2020 20th International Conference on Control, Automation and Systems, Busan, Korea, 13–16 October 2020; pp. 1151–1154.
14. Zhao, T.; Li, S.; Ngan, K.N.; Wu, F. 3-D reconstruction of human body shape from a single commodity depth camera. *IEEE Trans. Multimed.* **2019**, *21*, 114–123. [\[CrossRef\]](#)
15. Pires, I.M.; Hussain, F.; Marques, G.; Garcia, N.M. Comparison of machine learning techniques for the identification of human activities from inertial sensors available in a mobile device after the application of data imputation techniques. *Comput. Biol. Med.* **2021**, *135*, 104638. [\[CrossRef\]](#)
16. Beshara, P.; Chen, J.F.; Read, A.C.; Lagadec, P.; Wang, T.; Walsh, W.R. The reliability and validity of wearable inertial sensors coupled with the microsoft kinect to measure shoulder range-of-motion. *Sensors* **2020**, *20*, 7238. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Yu, W.; Zhu, M.; Wang, S. Research on the recognition algorithm of body posture ELM registration model based on Kinect. In Proceedings of the 2021 40th Chinese Control Conference, Shanghai, China, 26–28 July 2021; pp. 7318–7324.
18. Liu, W.; Li, M.; Liu, F.; Xu, Y. Dual predictive quaternion Kalman filter and its application in seamless wireless mobile human lower limb posture tracking. *Mob. Netw. Appl.* **2023**. [\[CrossRef\]](#)
19. Huang, Y.; Zhang, Y.; Xu, B.; Wu, Z.; Chambers, J.A. A new adaptive extended Kalman filter for cooperative localization. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *68*, 8671–8682. [\[CrossRef\]](#)
20. Cui, B.; Wei, X.; Chen, X.; Li, J.; Li, L. On sigma-point update of cubature Kalman filter for GNSS/INS under GNSS-challenged environment. *IEEE Trans. Veh. Technol.* **2019**, *68*, 8671–8682. [\[CrossRef\]](#)
21. Zhao, S.; Huang, B. Trial-and-error or avoiding a guess? Initialization of the Kalman filter. *Automatica* **2020**, *121*, 109184. [\[CrossRef\]](#)
22. Shen, C.; Zhang, Y.; Guo, X.; Chen, X.; Liu, J. Seamless GPS/inertial navigation system based on self-learning square-root cubature Kalman filter. *IEEE Trans. Ind. Electron.* **2021**, *68*, 499–508. [\[CrossRef\]](#)
23. Abbasi, J.; Salarieh, H.; Alasty, A. A motion capture algorithm based on inertia-Kinect sensors for lower body elements and step length estimation. *Biomed. Signal Process. Control.* **2021**, *64*, 102290. [\[CrossRef\]](#)
24. Kim, S.; Nozaki, T.; Murakami, T. An approach to categorization analysis for human motion by Kinect and IMU. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 6158–6162.
25. Bi, S.; Ma, L.; Shen, T.; Xu, Y.; Li, F. Neural network assisted Kalman filter for INS/UWB integrated seamless quadrotor localization. *PeerJ Comput. Sci.* **2021**, *7*, e630. [\[CrossRef\]](#)
26. Chen, M.; Li, Y.; Luo, X.; Wang, W.; Wang, L.; Zhao, W. A novel human activity recognition scheme for smart health using multilayer extreme learning machine. *IEEE Internet Things J.* **2018**, *6*, 1410–1418. [\[CrossRef\]](#)
27. Seenath, S.; Dharmaraj, M. Conformer-based human activity recognition using inertial measurement units. *Sensors* **2023**, *23*, 7357. [\[CrossRef\]](#)
28. Choi, H.; Jeon, H.; Noh, D.; Kim, T.; Lee, D. Hand-guiding gesture-based telemanipulation with the gesture mode classification and state estimation using wearable IMU sensors. *Mathematics* **2023**, *11*, 3514. [\[CrossRef\]](#)
29. Dahl, K.D.; Dunford, K.M.; Wilson, S.A.; Turnbull, T.L.; Tashman, S. Wearable sensor validation of sports-related movements for the lower extremity and trunk. *Med. Eng. Phys.* **2020**, *84*, 144–150. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Bijalwan, V.; Semwal, V.B.; Mandal, T.K. Fusion of multi-sensor-based biomechanical gait analysis using vision and wearable sensor. *IEEE Sens. J.* **2021**, *21*, 14213–14220. [\[CrossRef\]](#)
31. Yi, C.; Wei, B.; Ding, Z.; Yang, C.; Chen, Z.; Jiang, F. A self-aligned method of IMU-based 3-DoF lower-limb joint angle estimation. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–10. [\[CrossRef\]](#)
32. Zhang, J.; Li, P.; Zhu, T.; Zhang, W.A.; Liu, S. Human motion capture based on kinect and imus and its application to human-robot collaboration. In Proceedings of the 2020 5th International Conference on Advanced Robotics and Mechatronics, Shenzhen, China, 18–21 December 2020; pp. 392–397.
33. Fan, B.; Li, Q.; Tan, T.; Kang, P.; Shull, P.B. Effects of IMU sensor-to-segment misalignment and orientation error on 3-D knee joint angle estimation. *IEEE Sens. J.* **2021**, *22*, 2543–2552. [\[CrossRef\]](#)
34. Chen, L.; Yan, X.; Hu, D. A deep learning control strategy of IMU-based joint angle estimation for hip power-assisted swimming exoskeleton. *IEEE Sens. J.* **2023**, *23*, 15058–15070. [\[CrossRef\]](#)
35. Huttner, F.; Kalkkuhl, J.; Reger, J. Offset and misalignment estimation for the online calibration of an MEMS-IMU using FIR-filter modulating functions. In Proceedings of the 2018 IEEE Conference on Control Technology and Applications, Copenhagen, Denmark, 21–24 August 2018; pp. 1427–1433.
36. Rasoulzadeh, R.; Shahri, A.M. Accuracy improvement of a multi-MEMS inertial measurement unit by using an iterative UFIR filter. In Proceedings of the 2017 European Navigation Conference, Lausanne, Switzerland, 9–12 May 2017; pp. 279–286.
37. Sun, M.; Wang, Y.; Joseph, W.; Plets, D. Indoor localization using mind evolutionary algorithm-based geomagnetic positioning and smartphone IMU sensors. *IEEE Sens. J.* **2022**, *22*, 7130–7141. [\[CrossRef\]](#)
38. He, H.; Liu, G.; Zhu, X.; He, L.; Tian, G. Interacting multiple model-based human pose estimation using a distributed 3D camera network. *IEEE Sens. J.* **2019**, *19*, 10584–10590. [\[CrossRef\]](#)

39. Liu, H.; Stoll, N.; Junginger, S.; Zhang, J.; Ghandour, M.; Thurow, K. Human-mobile robot interaction in laboratories using Kinect sensor and ELM based face feature recognition. In Proceedings of the 2016 9th International Conference on Human System Interactions, Portsmouth, UK, 6–8 July 2016; pp. 197–202.
40. Akbari, A.; Thomas, X.; Jafari, R. Automatic noise estimation and context-enhanced data fusion of IMU and Kinect for human motion measurement. In Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks, Eindhoven, The Netherlands, 9–12 May 2017; pp. 178–182.
41. Cho, H.; Yeon, S.; Choi, H.; Doh, N.L. 3D pose estimation with one plane correspondence using kinect and IMU. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, Hamburg, Germany, 28 September–2 October 2015; pp. 1970–1975.
42. Xu, Y.; Wan, D.; Bi, S.; Guo, H.; Zhuang, Y. Predictive mode-ELM integrated assisted FIR filter for UWB robot localization. *Satell. Navig.* **2023**, *4*, 2. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.