mathematics

MDPI

Article

Wearable Inertial Sensor-Based Hand-Guiding Gestures Recognition Method Robust to Significant Changes in the Body-Alignment of Subject

Haneul Jeon, Haegyeom Choi, Donghyeon Noh, Taeho Kim and Donghun Lee *

Mechanical Engineering Department, Soongsil University, Seoul 06978, Republic of Korea
* Correspondence: dhlee04@ssu.ac.kr

Abstract: The accuracy of the wearable inertia-measurement-unit (IMU)-sensor-based gesture recognition may be significantly affected by undesired changes in the body-fixed frame and the sensor-fixed frame according to the change in the subject and the sensor attachment. In this study, we proposed a novel wearable IMU-sensor-based hand-guiding gesture recognition method robust to significant changes in the subject's body alignment based on the floating body-fixed frame method and the bi-directional long short-term memory (bi-LSTM). Through comparative experimental studies with the other two methods, it was confirmed that aligning the sensor-fixed frame with the reference frame of the human body and updating the reference frame according to the change in the subject's body-heading direction helped improve the generalization performance of the gesture recognition model. As a result, the proposed floating body-fixed frame method showed a 91.7% test accuracy, confirming that it was appropriate for gesture recognition under significant changes in the subject's body alignment during gestures.

Keywords: gesture recognition; bi-directional LSTM; wearable sensor; biomechanics; hand-guiding gesture

MSC: 68T01; 68T05

Citation: Jeon, H.; Choi, H.; Noh, D.; Kim, T.; Lee, D. Wearable Inertial Sensor-Based Hand-Guiding Gestures Recognition Method Robust to Significant Changes in the Body-Alignment of Subject. *Mathematics* **2022**, *10*, 4753. https://doi.org/10.3390/math10244753

Academic Editor: Andrea Prati

Received: 7 November 2022
Accepted: 3 December 2022
Published: 14 December 2022

1. Introduction

Human gesture recognition technology has steadily been applied to healthcare [1] and remote control [2], along with the spread of compact small-sized mobile devices and the development of deep learning technology. In the industrial field, gesture recognition is also being used for remote control of various automation systems to prevent musculoskeletal diseases of workers [3,4].

Most gesture recognition methods are implemented with vision sensors or wearable sensors. The vision-based methods have mostly used RGB cameras [4–6] and Kinects [3]. In the study of Nuzzi et al. [5], five pieces of hand-gesture data collected through an RGB camera were used for accurate hand position and gesture recognition in RGB images using the R-CNN algorithm to 92.6%. Jiang et al. [6], using an RGB-D sensor, proposed a skeletonization algorithm for effective gesture recognition and classified 24 hand gestures collected with the Kinect with an accuracy of 93.63% through the CNN. However, in the case of using these RGB and RGB-D sensors, the gesture capture area is inevitably restricted to inside the sensor field of view (FOV), and the light reflections as well as low light can severely degrade the recognition accuracy [7].

The wearable-sensor-based gesture recognition methods have mainly used IMU sensors [8–14], electromyography sensors [15–18], and multimodal wearable devices [19]. Abualola et al. [8] proposed an IMU-integrated glove for hand-gesture recognition. The system tracks fine-grain hand movements using inertial and attitude measurements. Gestures

are recognized in real-time based on Linear Discriminant Analysis (LDA) with an accuracy of 85%. Suri et al. [9] studied sign language recognition using a wrist-worn IMU sensor, and 30 pieces of sign language sentence data obtained by the IMU were classified with an average accuracy of 94.6%. Moreover, Khassanov et al. [10,11] proposed multiple IMU-sensor-based methods for recognizing supervisory user interface commands, such as manipulator on/off and operation pause/resume for a mobile manipulator system.

However, because the accuracy of the wearable IMU-sensor-based gesture recognition may be significantly affected by undesired changes in the subject's heading direction and the sensor attachment's pose, the subjects should keep their initial body-alignment identified at the sensor calibration step to recognize the target gestures correctly, as shown in Figure 1. Moreover, there are no studies that explicitly report on these issues. Thus, in this study, we propose a new wearable IMU-sensor-based hand-guiding gesture recognition method robust to significant changes in the subject's body alignment with the floating body-fixed frame method and the bi-directional LSTM. Comparative experimental studies are also performed for the five hand-guiding gesture classifications and the five dumbbell exercise classifications to validate the proposed method's significance compared to the previous methods.

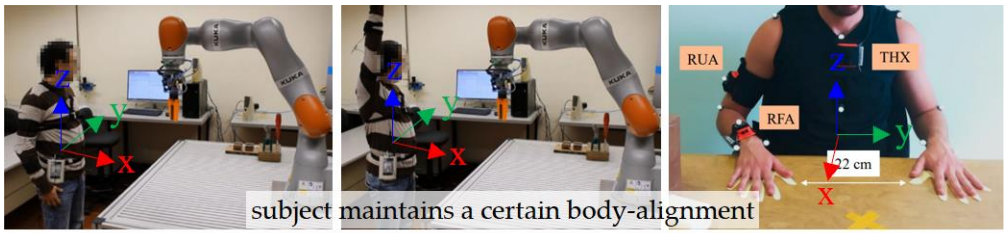


Figure 1. Wearable inertial-sensor-based remote manipulation; the subject maintains a certain body-alignment [12,13].

2. Problem Definition

As shown in Table 1 below, in this study, the wearable IMU-sensor-based dynamic gesture recognition method is classified into the following three according to the satisfaction of the key requirements in the wearable-sensor-based methods, such as the creation of a new body-fixed frame, the change in the sensor-fixed frame every time a wearable sensor is attached to the body, and sensor recalibration according to changes in the subject's body-heading direction. Method A [12,13] only performs the creation of a new body-fixed frame, and method B [20] performs the creation of a new the body-fixed frame after the creation of a new body fixed frame. To gain full insight into the need for method C proposed in this study, in this section, an in-depth data-based discussion is conducted on the two different hand-guiding gestures' data collected while applying intentional changes to the subject's body-heading direction.

Table 1. Classification of the motion recognition method according to the requirement satisfaction.

Requirement	*A*	*B*	*C*
Creating and referencing subject's body-fixed frame	○	○	○
Aligning all sensor-fixed frames equally	×	○	○
Floating body-fixed frame	×	×	○

○: method is used ×: method is not used.

2.1. Concepts of Creating Body-Fixed Frame and Aligning All Sensor-Fixed Frames Equally

Figure 2 explains the concept of the sensor calibration process to align all sensor-fixed frames in the same orientation as the body-fixed frame. The IMU sensor has three outputs: orientation, angular velocity, and acceleration: the orientation of the sensor is expressed as the relative position of the sensor with respect to the inertial frame, and the angular velocity and acceleration are expressed with respect to the sensor-fixed frame. Here, it

is a very self-evident fact that attaching the output reference frame of the inertial sensor to the subject's body rather than the globally fixed inertial frame will significantly help the realization of a consistent gesture recognition algorithm. However, despite the fact that the body-fixed frame generated based on the subject's body-heading direction provides a consistent measurement reference frame for time-varying body-heading directions each time, the low reproducibility of the sensor pose that occurs when the sensor is attached will still cause significantly high uncertainties in collecting consistent feature data from wearable IMU sensors. Then, the resetting all the sensor-fixed frames of all different sensors in the same way as the body-fixed frames created earlier will help collect sensor data consistently. As a result, it will also help improve the generalization performance of the gesture recognition model trained with those sensor data.

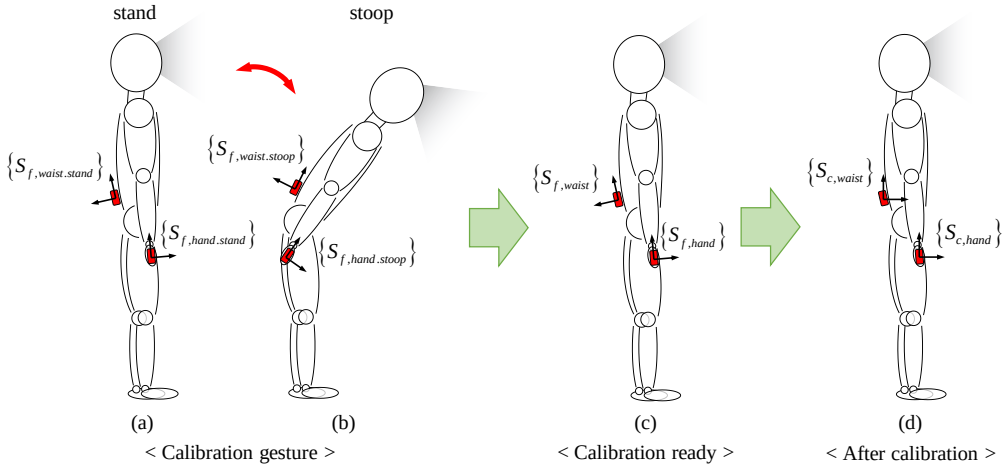


Figure 2. Overall procedure of the hand-mounted IMU sensor calibration for aligning all sensor-fixed frames equally with a waist-worn IMU sensor in this study: (**a**) standing, (**b**) stooping, (**c**) home pose or calibration-ready pose, and (**d**) measurement-ready pose after standing back.

2.2. Description about Concept of Floating Body-Fixed Frame Method

Let us consider a case in which a kernel closely related to direction, such as inward and outward turning in the hand-guiding gesture, exists in the target behaviors to be recognized. As shown in Figure 3, the subject's body alignment to the initially defined body-fixed frame can be time-varied according to the various application scenarios of the gesture recognition model. It can also be a factor that significantly hinders the generalization performance of the gesture recognition model because it is impossible to collect learning data for all labels in all possible heading directions of the subjects, for example, in the case of a gesture meaning facing forward and turning left or turning to the right, and the same gesture being taken by turning 180 degrees back, due to the reverse effect of feature data's patterns or the correlation of features with other labels, etc. A critical decrease in the recognition accuracy may occur, and a mode collapse problem of a specific label among the correlated labels may also occur.

Figure 3 shows how simple hand-guiding gestures are repeatedly performed under the general scenario where the subject's body alignment to the initially defined body-fixed frame is continuously time-varying. Figure 3a shows that the subject maintains their initial body alignment within a certain range with respect to the initially defined body-fixed frame, and Figure 3b,c show that the subject's body alignment is intentionally rotated about 90 and 180 degrees about the z-axis, respectively, compared to the initial state. On the right side, angular velocity and acceleration plots measured for each case for the same two (turning inward and turning outward) hand-guiding gestures are shown for comparison. As a result, in the case of angular velocity, it can be seen that the activation pattern for each case is different for a hand motion performed once. For example, in the case of Figure 3a,c, it can be seen that the phases of the x-axis components are inverted, unlike those in which the phases and amplitudes are very similar. In addition, comparing the turning inward

of Figure 3a and the turning outward of Figure 3c, the angular velocity components are almost the same. Let us consider comparing gestures that have a phase difference of 180 degrees with each other, such as turning inward and outward in Figure 3a,c. When turning inward is performed while the subject's body alignment to the initially defined body-fixed frame is matched, the user's body rotates 180 degrees about the z-axis with respect to the initially defined body-fixed frame. When turning inward, the phases of the angular velocity components, which are one of the key features, are reversed, so the gesture recognition model may incorrectly predict it as turning outward. Based on the above visual inspections on the feature data measured in different body-heading directions, we can conclude that the body's misalignment to the initial body alignment causes severe degradation in recognition accuracy due to the uncertainties in distinguishing labels. Therefore, in the next section, the sensor calibration and floating body-fixed frame methods are described in detail. In Section 4, the performance differences in multi-class dynamic hand-gesture recognition for the three methods in Table 1 are experimentally compared and discussed.

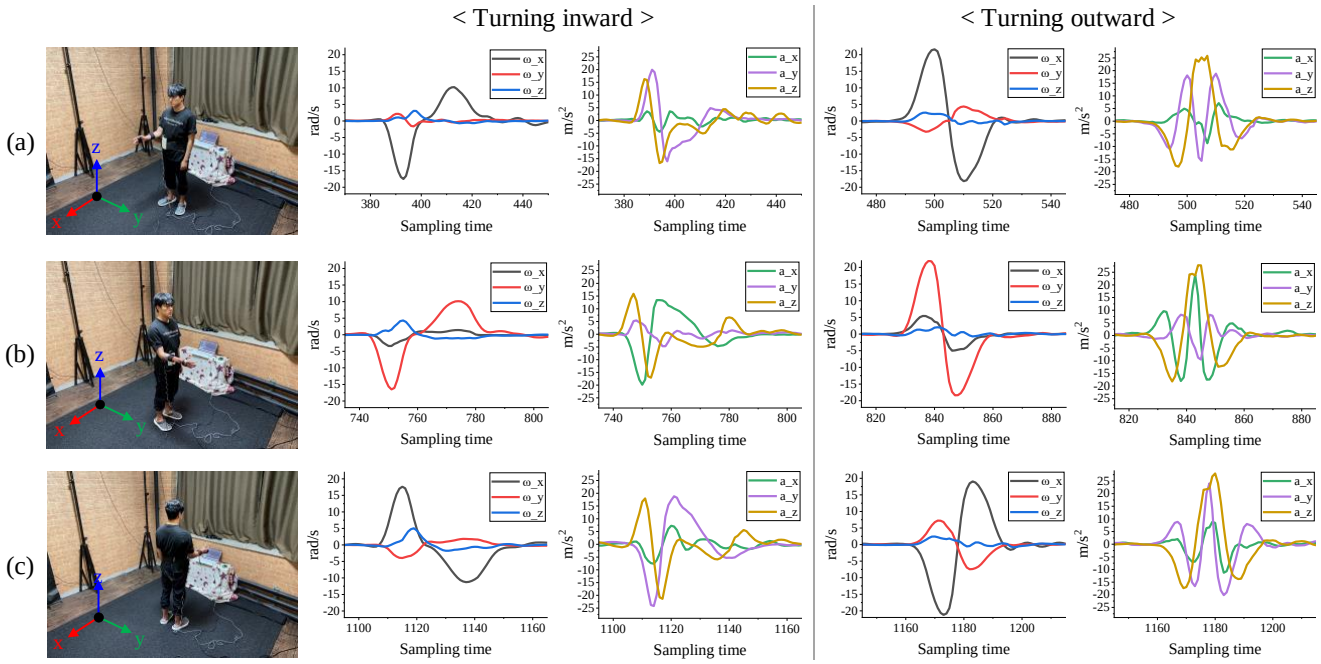


Figure 3. Comparison of angular velocity and acceleration of the back of the hand measured while simple handshaking in three different body-heading directions according to rotation about the z-axis: (**a**) initial body alignment, (**b**) 90 degrees, and (**c**) 180 degrees.

3. Method

3.1. Sensor Calibration

This section describes the protocol of generating the body-fixed frame $\{B_f\}$ through the calibration gesture of standing-stooping shown in Figure 2. In biomechanics [21,22], most of the motions of the human body are analyzed with respect to the three mutually orthogonal motion planes (sagittal plane, frontal plane, transverse plane), so it would be a very reasonable choice to define the body-fixed frame by aligning it with the principal axes of these motion planes.

Figure 2 shows the entire process of resetting the orientations of different sensors equally by creating a body-fixed frame $\{B_f\}$ aligned with the principal axes of the motion planes of the human body. Figure 2a,b show the subject's calibration gesture to create $\{B_f\}$ in order. The average orientation at each stand and stoop calibration gesture can be calculated from the set of $\{S_{f,waist}\}$ with respect to the inertial frame collected for about 5 s at a sampling rate of 100 Hz [20]. Moreover, Figure 2c shows a calibration-ready or home pose to match the orientations of $\{B_f\}$ and $\{S_{f,j}\}$ created earlier. The subject puts their arms

to the sides of both thighs with the back of their hand facing outward. After the subject takes a calibration-ready pose, 5 s later, the orientation of $\{S_{f,j}\}$ is reset to $\{S_{c,j}\}$, which is the same orientation as $\{B_f\}$. The axis (vector $\boldsymbol{K}$) for the transition from the average stand to the stooping pose can be obtained through Equation (1).

$$\begin{gathered} {}^{G}_{k}R = {}^{G}_{f,waist,stoop}R^{T} \cdot {}^{G}_{f,waist,stand}R^{T} \\ \theta = \cos^{-1}\left(\frac{trace({}^{G}_{k}R)-1}{2}\right), \;\; K = \frac{1}{2\sin\theta} \cdot \begin{bmatrix} r_{32}-r_{23} \\ r_{13}-r_{31} \\ r_{21}-r_{12} \end{bmatrix} \end{gathered} \tag{1}$$

Through Algorithm 1 below, $\{B_f\}$ can be defined in the form of SO(3) as follows.

$${}^{G}_{B_f}R = \begin{bmatrix} x_{B_f} & y_{B_f} & z_{B_f} \end{bmatrix} \tag{2}$$

where z_{B_f} denotes a unit length vertical vector opposite to gravity of $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$, y_{B_f} denotes the subject's transverse axis, and x_{B_f} denotes the axis representing the subject's body-heading direction obtained through the cross-product of $\hat{k}$ and z_{B_f}. Algorithm 1 describes a detailed protocol for generating a body-fixed frame based on the following three assumptions:

(1) $\hat{k}$ coincides with the subject's transverse axis;
(2) z_{B_f} axis coincides with the subject's longitudinal axis;
(3) x_{B_f} axis is aligned with the subject's anteroposterior axis.

Algorithm 1 *Create body-fixed frame*

procedure sensor data $\left({}^{G}_{S_{f,j}}R, {}^{G}\vec{a}_{S_{f,j}}, {}^{G}\vec{\omega}_{S_{f,j}}\right)$
  While *about* 5 s
    Maintain standing posture
    Save orientation data ${}^{G}_{S_{f,j}}R$
  end
  Update ${}^{G}_{f,j.stand}R \leftarrow avg(saved\ orientation\ data)$
  While 5 *s*
    Maintain stooping posture
    Save orientation data ${}^{G}_{S_{f,j}}R$
  end
  Update ${}^{G}_{f,j.stoop}R \leftarrow avg(saved\ orientation\ data)$
  Calculate *vector* $\boldsymbol{k} \leftarrow \left({}^{G}_{f,j.stoop}R^{T} \cdot {}^{G}_{f,j.stand}R\right)$
  Calculate *vector* $\boldsymbol{x} \leftarrow \left(vector\ \ \mathrm{k} \times \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T\right)$
  Return *Body- fixed frame* $\leftarrow \begin{bmatrix} x_{B_f} & k_{B_f} & z_{B_f} \end{bmatrix}$

3.1.1. Orientation w.r.t Body-Fixed Frame

In this part, a method of changing the orientation of the sensor expressed for the inertial frame $\{G\}$ to be expressed with respect to $\{B_f\}$ and, at the same time, resetting the sensor-fixed frame $\{S_j\}$ having different orientations to be the same as $\{B_f\}$ is described. First, at the calibration-ready pose in Figure 2c, the relative orientation of the current sensor-fixed frame of each sensor with respect to $\{B_f\}$ is obtained through Equation (3), and then, the new sensor-fixed frame $\{C_j\}$ of each sensor initially identical to $\{B_f\}$ is created.

$${}^{S_{f,j.stand}}_{C_j}R = {}^{G}_{S_{f,j.stand}}R^{T} \cdot {}^{G}_{B_f}R \tag{3}$$

As a result, the orientation of $\{C_j\}$ with respect to $\{B_f\}$ can be calculated as in Equation (4).

$${}^{B_f}_{C_{,j}}R = {}^{G}_{B_f}R^{T} \cdot {}^{G}_{S_{f,j}}R \cdot {}^{S_{f,j.stand}}_{C_{,j}}R \tag{4}$$

3.1.2. Angular Velocity and Acceleration w.r.t Body-Fixed Frame

In this part, the angular velocity and acceleration expressed with respect to the original sensor-fixed frame $\{S_{f,j}\}$ of each sensor transform into an expression with respect to frame $\{B_f\}$ through Equation (5) as in the orientation of I.

$$\begin{aligned} {}^{B_f}\vec{a}_{S_{f,hand}} &= {}^{G}_{B_f}R^T \cdot {}^{G}_{S_{f,hand}}R \cdot \left({}^{S_{f,hand}}\vec{a}\right) \\ {}^{B_f}\vec{\omega}_{S_{f,hand}} &= {}^{G}_{B_f}R^T \cdot {}^{G}_{S_{f,hand}}R \cdot \left({}^{S_{f,hand}}\vec{\omega}\right) \end{aligned} \tag{5}$$

The contents of I and II are summarized as in Algorithm 2 below.

Algorithm 2 *Align all sensor fixed frames equally*

procedure sensor data, body fixed frame, orientation of the initial posture
Calculate *rotated frames mapping*
${}^{S_{f,j.stand}}_{C_j}R = {}^{G}_{S_{f,j.stand}}R^T \cdot {}^{G}_{B_f}R$
Calculate *sensor orientation w.r.t* $\left\{B_f\right\}$
${}^{B_f}_{C_j}R = {}^{G}_{B_f}R^T \cdot {}^{G}_{S_{f,j}}R \cdot {}^{S_{f,j.stand}}_{C_j}R$
Calculate *sensor acceleration w.r.t* $\left\{B_f\right\}$
${}^{B_f}\vec{a}_{S_{f,hand}} = {}^{G}_{B_f}R^T \cdot {}^{G}_{S_{f,hand}}R \cdot {}^{G}\vec{a}_{S_{f,hand}}$
Calculate *sensor rate of turn w.r.t*
${}^{B_f}\vec{\omega}_{S_{f,hand}} = {}^{G}_{B_f}R^T \cdot {}^{G}_{S_{f,hand}}R \cdot {}^{G}\vec{\omega}_{S_{f,hand}}$
Return *sensor data w.r.t* $\left\{B_f\right\}$

3.2. Floating Body-Fixed Frame

As mentioned in the problem definition part, the subject's body misalignment against the initial body-fixed frame $\{B_f\}$ will increase the uncertainties in the performance of the gesture recognition model. Therefore, in this section, a floating-body-fixed frame $\{FB_f\}$ that can continuously recalibrate body alignment based on the subject's time-varying body-heading direction information is obtained through Equation (6).

$${}^{G}_{FB_f}R = {}^{G}_{S_{f,waist}}R \cdot {}^{G}_{S_{f,waist,stand}}R^T \cdot {}^{G}_{B_f}R \tag{6}$$

Then, Equation (7) is used to update the orientation of the new sensor-fixed frame $\{C_j\}$ of each sensor with the expression with respect to frame $\{FB_f\}$ updated in real time.

$${}^{FB_f}_{C_j}R = {}^{G}_{FB_f}R^T \cdot {}^{G}_{S_{f,j}}R \cdot {}^{S_{f,j.stand}}_{C_j}R \tag{7}$$

Then, Equation (8) is used to transform the angular velocity and acceleration with respect to the original sensor-fixed frame $\{S_{f,j}\}$ of each sensor into the expression with respect to frame $\{FB_f\}$. All the updating of the body-fixed frame according to change in time-varying body-alignment is summarized in Algorithm 3.

$$\begin{aligned} {}^{FB_f}\vec{a}_{S_{f,hand}} &= {}^{G}_{FB_f}R^T \cdot {}^{G}_{S_{f,hand}}R \cdot \left({}^{S_{f,hand}}\vec{a}\right) \\ {}^{FB_f}\vec{\omega}_{S_{f,hand}} &= {}^{G}_{FB_f}R^T \cdot {}^{G}_{S_{f,hand}}R \cdot \left({}^{S_{f,hand}}\vec{\omega}\right) \end{aligned} \tag{8}$$

Algorithm 3 *Updating the body-fixed frame according to change in subject's time-varying body-alignment*

procedure sensor data, orientation of the initial posture
Calculate *floating body fixed frame*
$${}^{G}_{FB_f}R = {}^{G}_{S_{f,waist}}R \cdot {}^{G}_{S_{f,waist,stand}}R$$
Calculate *sensor orientation w.r.t* $\left\{FB_f\right\}$
$${}^{FB_f}_{C_{,j}}R = {}^{G}_{FB_f}R^{T} \cdot {}^{G}_{S_{f,j}}R \cdot {}^{S_{f,j,stand}}_{C_{,j}}R$$
Calculate *sensor acceleration w.r.t* $\left\{FB_f\right\}$
$${}^{FB_f}\vec{a}_{S_{f,hand}} = {}^{G}_{FB_f}R^{T} \cdot {}^{G}_{S_{f,hand}}R \cdot {}^{G}\vec{a}_{S_{f,hand}}$$
Calculate *sensor rate of turn w.r.t* $\left\{FB_f\right\}$
$${}^{FB_f}\vec{\omega}_{S_{f,hand}} = {}^{G}_{FB_f}R^{T} \cdot {}^{G}_{S_{f,hand}}R \cdot {}^{G}\vec{\omega}_{S_{f,hand}}$$
Return *sensor data w.r.t* $\left\{FB_f\right\}$

3.3. Practical Application to the Multi-Class Classification of the Hand-Guiding Gestures

In this study, five gesture modes were defined: shaking inward (*si*), shaking outward (*so*), turning inward (*ti*), and turning outward (*to*), indicating not only via motion (*vm*), which means unintentional guiding gesture, but also a combination of linear/angular and inward/outward, as shown in Figure 4. This mode classification problem of the dynamic hand-guiding gesture will be affected not only in instantaneous mode at every moment but also in long-term dependency between these instantaneous modes due to the complex and non-linear dynamic characteristics of hand-guiding motion itself.

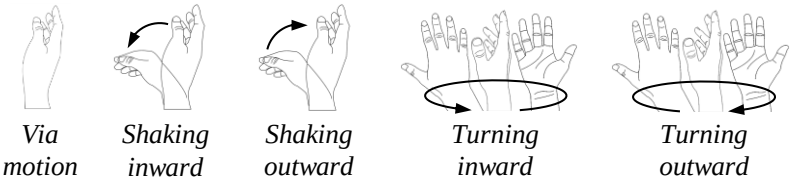


Figure 4. Illustration of five hand-guiding gestures to be classified.

Therefore, sufficient biomechanical and ergonomic insight for the corresponding gestures should be necessary to select a classification method suitable for these human gesture classification problems. For example, it can be seen from Figure 4 that even when intentional *so* (*shaking outward*) is performed, unintentional *si* (*shaking inward*) is always performed in pairs, and vice versa. As shown in Figure 5, this phenomenon can be identified through visual inspection of ω_z among the shaking outward and inward motion data collected from 6 subjects. In the case of shaking outward and inward, it can be seen that they show very similar behavior to each other, except that they show a phase difference of 180 degrees. Here, if we consider repetitive shaking inward as in Figure 5a, the prediction model based on the sliding window method will predict shaking inward alternately inward and outward.

Similarly, in the case of *so*, the classification model will internally classify by repeatedly switching instantaneous modes of *so* and *si* when classifying the so mode. In order to recognize such a continuous gesture, it is very important to accurately recognize the gesture's intention by identifying its start moment. In particular, in the case of a classification model using unidirectional classification, it will be difficult to distinguish the start moment of the corresponding gesture due to the model's limited memory as the gestures become longer. In addition, in the case of intentional outward (inward) shaking, as shown in Figure 5, it can be seen that the intensity of inward (outward) shaking, which is a returning gesture, and the intensity and time interval of the intended gesture are different from each other. In general, most subjects made the intentional gestures stronger and faster when they took the hand-guiding gesture. In the case of intensity, it was confirmed that the intensity of the returning gestures was about 27~45% lower than that of the intended gesture, and the time interval of the intended gesture compared to the returning gesture was about

26~44% lower. Based on these biomechanical and ergonomic insights, the bi-directional LSTM [23], which is capable of instantaneous mode classification reflecting the temporal context even at the last time point by giving backward feedback of the classification results at the present time in the direction of the beginning of the time horizon, should be the most suitable gesture recognition algorithm for this study.

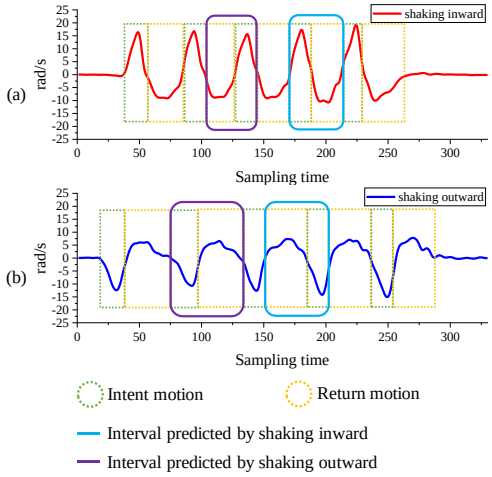


Figure 5. Plot of angular velocity about z-axis in (**a**) shaking inward and (**b**) outward motion data collected from 6 subjects.

Figure 6 shows the overall framework of the bi-directional LSTM-based five hand-guiding gesture classifications. In order to understand not only the instantaneous mode but also the long-term dependencies between the instantaneous modes, nine features (three-dimensional orientation, acceleration, and angular velocity with respect to the frame $\{FB_f\}$) collected from the IMU sensor are captured at every 10 ms for a predefined time-horizon length (400 ms in this case). Then, this way, 9×40 two-dimensional input data collected for 400 ms at a sampling rate of 10 ms are input to a bi-directional LSTM after the normalization process. Finally, the input 9×40 data are transformed into the 5×1 probability vector by activating the 9×1 hidden vector as output after identifying the gesture context through the forward LSTM model and the backward LSTM model in this study.

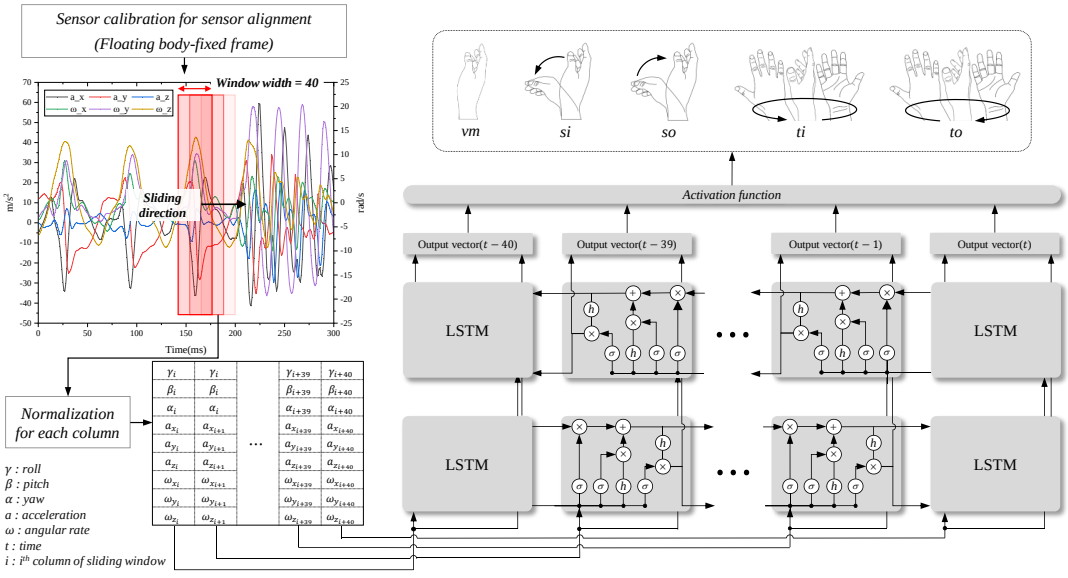


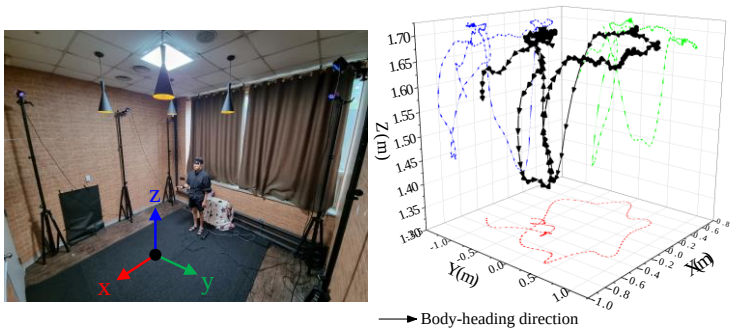
Figure 6. Overall framework of the bi-directional LSTM-based different hand-guiding gesture classification.

4. Experiment and Discussion

This section describes the overall data collection, data expansion, and learning and evaluation results for learning and evaluating the hand-guiding gesture recognition model in Figure 6.

4.1. Experimental Setup

In this study, six subjects collected the gesture dataset of hand-guiding gestures represented in Figure 7 using Xsens' MTx wearable inertial sensor in the environment shown in Figure 7a below. After configuring the capture volume of 300 × 300 × 250 mm^3 with six prime-13 vision sensors and Optitrack's motion capture system, the subjects' walking trajectory and body-heading directions were recorded and visualized to prove the appropriateness of the training dataset acquisition process. During dataset acquisition, the subjects walked freely within the capture volume and repeatedly performed hand-guiding gestures, and the changes in the subjects' pose and body-heading direction are shown in Figure 7b.



(a) (b)

Figure 7. (**a**) Experimental environment including Optitrack's six prime-13 vision sensors and Xsens wearable IMU sensors, and changes in (**b**) pose and body-heading direction of subjects in the data acquisitions.

4.2. Training and Test Dataset Acqusition

In addition, to prove that gesture recognition based on the floating body-fixed frame method is more robust to significant changes in the motion and body-alignment of the subject than methods A and B, the gesture data for methods A and B were also collected. As a result, about 126,165 training datasets and 84,592 validation datasets were collected from 4 subjects, and 52,259 test datasets were collected from 2 other subjects with the same protocol as the training data.

4.3. Training, Test, and Result Discussion

The learning condition of the model was batch size = 5000, epoch = 1000, and learning rate = 0.001. In addition, to prevent convergence to the local minima as learning progresses, the learning rate was lowered with the callback function to allow it to escape from the local minima. In addition, to prevent overfitting, learning was stopped when the performance of test loss was no longer improving using the early stopping method, and the time horizon length closely related to the temporal context was set to 400 ms, according to the results presented in Table 2. For preliminary validation of the bi-directional LSTM selected as the classification model for this study, we first compared and validated with the RNN (a.k.a vanilla RNN) [24] and LSTM [25]. As a result, the training and test accuracies for each label of RNN, LSTM, and bi-directional LSTM based on the floating body-fixed frame method are presented in Table 3 for comparison purposes. Table 4 also shows that the proposed floating body-fixed frame method has a significant performance improvement compared to the other two methods in hand-guiding gesture recognition under significant changes in the subject's body alignment based on the bi-directional LSTM model.

Table 2. Comparison of training and test accuracy of bi-directional LSTM model for the floating body-fixed frame method according to the change in time-horizon length.

-	*Training/Test Accuracy by Label* (%)					
	vm	***si***	***So***	***ti***	***to***	***Total***
300 ms	99.3/71.9	99.7/83.5	99.8/99.9	99.3/91.6	99.6/82.2	99.5/84.0
400 ms	97.7/89.6	97.0/92.6	96.4/89.0	99.3/98.7	98.9/99.4	97.9/91.7
500 ms	99.9/57.6	99.6/96.1	99.8/99.8	99.6/75.6	99.8/90.4	99.7/82.6
600 ms	99.9/70.0	100/97.4	100/63.1	99.9/81.4	100/70.3	99.9/72.9

Table 3. Comparison of training and test accuracy of RNN and LSTM and bi-directional LSTM model for the 5 different hand-guiding gesture classifications.

-	*Training/Test Accuracy by Label* (%)					
	vm	***si***	***so***	***ti***	***to***	***Total***
Vanilla RNN	91.8/88.6	81.8/59.5	78.8/62.1	90.7/83.9	78.4/61.7	83.6/66.4
Vanilla LSTM	98.0/89.1	98.2/90.2	94.4/71.2	98.1/85.3	98.7/98.9	97.5/85.1
Bi-directional LSTM	97.7/89.6	97.0/92.6	96.4/89.0	99.3/98.7	98.9/99.4	97.9/91.7

Table 4. Comparison of training and test accuracy of the floating body-fixed frame method over methods A and B with the bi-directional LSTM model.

-	*Training/Test Accuracy by Label* (%)					
	vm	***si***	***so***	***ti***	***to***	***Total***
Method A	98.4/49.6	92.8/37.9	93.6/43.4	98.4/95.0	98.5/99.6	96.3/57.7
Method B	99.4/77.5	97.1/64.7	99.2/75.8	99.7/82.6	99.8/75.8	99.0/74.6
Method C	97.7/89.6	97.0/92.6	96.4/89.0	99.3/98.7	98.9/99.4	97.9/91.7

In addition, for preliminary validation of the bi-directional LSTM selected as the classification model for this study, the training and test accuracy results for each label of RNN, LSTM, and bi-directional LSTM based on the floating body-fixed frame method are presented in Table 3 for comparison purposes. Table 4 also shows that the proposed floating body-fixed frame method has a significant performance improvement compared to the other two methods in hand-guiding gesture recognition under significant changes in the subject's body alignment based on the bi-directional LSTM model.

As shown in Table 3, the test accuracy of the bi-directional LSTM was highest at 91.7%, followed by the vanilla LSTM with 85.1% and the vanilla RNN with the lowest at 66.4%. In the case of LSTM, it is often wrong to classify *si* as *so* and vice versa, because it is difficult to correctly determine the gesture context with unidirectional LSTM. In particular, in the case of RNN, there was a considerably high frequency of misclassifying *si* into *so* and *to* as *ti*, which is interpreted to be caused by a decrease in classification accuracy due to long-term dependency when the gesture length is increased.

Table 4 shows the comparison result of training and test accuracy of the floating body-fixed frame method over methods A and B using 400 ms bi-directional LSTM. It can be seen that method C showed a test accuracy of 91.7%, and method B, which only aligns the sensor-fixed frame to the body-fixed frame, showed a test accuracy of 74.6%. That is, method A and method B decreased the test accuracy by −31.0% and −18.6%, respectively, compared to method C, which means that method C was more robust to the significant changes in the body alignment than other methods as claimed in the problem definition section. In addition, the lowest test accuracy of method A of 57.7% proves that the low reproducibility in the attached pose of the wearable sensor to the body segments causes significantly high uncertainties in collecting consistent feature data of the hand-guiding gestures. In summary, through this study, it was experimentally confirmed that aligning the sensor-fixed frame with the reference frame of the human body and updating the reference frame according to the change in the subject's body-heading direction helped improve the generalization

performance of the gesture recognition model. The confusion matrix of the test set using the floating body-fixed frame and bi-directional LSTM model proposed in this study is shown in Figure 8.

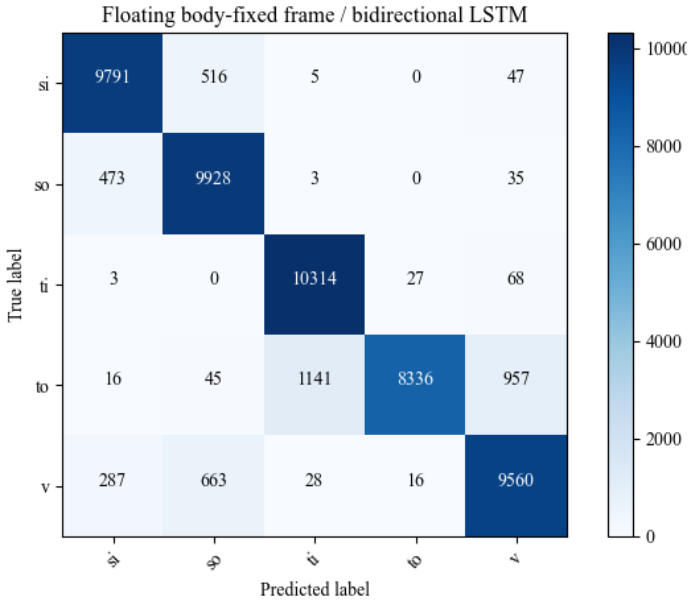


Figure 8. Resultant confusion matrix for the hand-guiding gesture recognition realized through the bi-directional LSTM model.

5. Conclusions

In this study, we proposed a wearable IMU-sensor-based hand-guiding gesture recognition method robust to significant changes in the motion and body-alignment of the subject based on the floating body-fixed frame method together with the bi-directional LSTM. To validate the research contributions of the proposed method, method A and method B were chosen according to the satisfaction of the key requirements in the wearable sensor-based methods, such as the creation of a new body-fixed frame and the change in the sensor-fixed frame in sensor attachment to the body. As a result, as shown in Table 2, it was confirmed that an excellent average test accuracy of 91.7% was achieved even under the condition of applying an intentional change to the subject's body-heading direction. For validating the significance of the proposed method on the gestures other than forearm-oriented hand-guiding gestures covered in this study, the additional validation study was conducted on four dumbbell exercises that simultaneously use both the arm and forearm, as shown in Figure 9. About 94,605 training datasets and 63,104 validation datasets were collected from 3 subjects, and 105,401 test datasets were collected from 2 other subjects with the same protocol as the training data. The training and test accuracies represented in Table 5 were obtained through three repetitive model trainings for each method. According to the results, the test accuracy of method C still showed significant improvement by 45.2% and 14.4%, respectively, compared to methods A and B, as with the hand-guiding gesture case in Table 5. Therefore, it can be seen that the FBF method could improve generalization performance for forearm-oriented hand-guiding gestures and general gestures using the entire upper arm part. In future work, we plan to apply the method proposed and verified through this study to the gesture-recognition-based remote control of mobile manipulators.

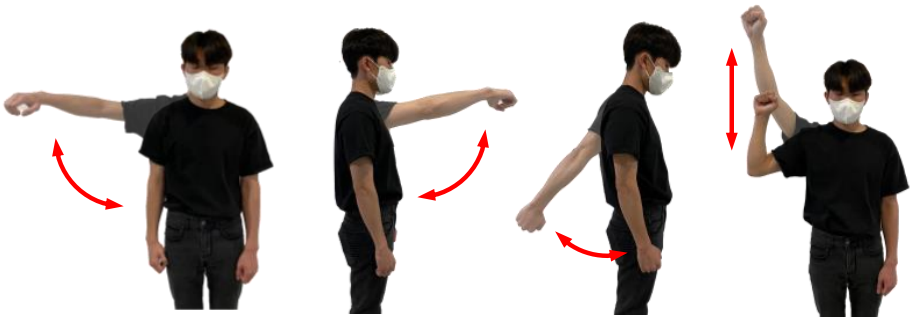


Figure 9. Dumbbell exercises: (**a**) side lateral raise (SLR), (**b**) front raise (FR), (**c**) dumbbell kickback (DKB), and (**d**) shoulder press.

Table 5. Comparison of training and test accuracy of the floating body-fixed frame method over methods A and B for the dumbbell exercises.

-	*Training/Test Accuracy by Label* (%)					
	SLR	***FR***	***DKB***	***SP***	***VM***	***Total***
Method A	95.6/74.0	94.5/72.8	97.7/60.9	97.9/76.1	96.5/37.6	96.4/58.0
Method B	98.6/74.2	98.2/73.8	99.5/77.4	99.9/97.4	99.5/54.7	99.0/73.6
Method C	99.7/77.9	99.6/92.9	99.9/85.9	99.9/96.2	99.5/71.9	99.7/84.2

Author Contributions: Conceptualization, D.L. and H.J.; methodology, D.L., H.J. and H.C.; software, H.J. and H.C.; validation, H.J., D.N. and T.K.; formal analysis, H.J., H.C. and D.N.; data curation, H.J., H.C., D.N. and T.K.; writing—original draft preparation, H.J. and D.L.; writing—review and editing, H.J. and D.L.; visualization, H.J.; supervision, H.J. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022 R1F1A1074704); the MSIT (Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2022-RS-2022-00156360) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation); Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00218); Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (N000P0017033).

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

Nomenclature

R	Rotation matrix
$\{G\}$	Global reference frame
$\{B_f\}$	Body-fixed frame
$\{FB_f\}$	Floating body-fixed frame
$\{S_{f,j}\}$	Sensor-fixed frame of j^{th} IMU sensor
$\{C_j\}$	Calibrated sensor-fixed frame
$\{S_{f,j.stand}\}$	Sensor-fixed frame at initial standing posture
$\{S_{f,j.stoop}\}$	Sensor-fixed frame at initial stooping posture
$\vec{a}$	Acceleration
$\vec{\omega}$	Angular rate
SO(3)	Three-dimensional orthogonal group

References

1. Al-Hammadi, M.; Muhammad, G.; Abdul, W.; Alsulaiman, M.; Bencherif, M.A.; Mekhtiche, M.A. Hand gesture recognition for sign language using 3DCNN. *IEEE Access* **2020**, *8*, 79491–79509. [CrossRef]
2. Chen, B.; Hua, C.; Dai, B.; He, Y.; Han, J. Online control programming algorithm for human–robot interaction system with a novel real-time human gesture recognition method. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881419861764. [CrossRef]
3. Popov, V.; Ahmed, S.; Shakev, N.; Topalov, A. Gesture-based Interface for Real-time Control of a Mitsubishi SCARA Robot Manipulator. *IFAC-PapersOnLine* **2019**, *52*, 180–185. [CrossRef]
4. Chen, J.; Ji, Z.; Niu, H.; Setchi, R.; Yang, C. An auto-correction teleoperation method for a mobile manipulator using gaze tracking and hand motion detection. In Proceedings of the Annual Conference Towards Autonomous Robotic Systems, London, UK, 3–5 July 2019; pp. 422–433.
5. Nuzzi, C.; Pasinetti, S.; Lancini, M.; Docchio, F.; Sansoni, G. Deep learning-based hand gesture recognition for collaborative robots. *IEEE Instrum. Meas. Mag.* **2019**, *22*, 44–51. [CrossRef]
6. Jiang, D.; Li, G.; Sun, Y.; Kong, J.; Tao, B. Gesture recognition based on skeletonization algorithm and CNN with ASL database. *Multimed. Tools Appl.* **2019**, *78*, 29953–29970. [CrossRef]
7. Suarez, J.; Murphy, R.R. Hand gesture recognition with depth images: A review. In Proceedings of the 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot And Human Interactive Communication, Paris, France, 9–13 September 2012; pp. 411–417.
8. Abualola, H.; Al Ghothani, H.; Eddin, A.N.; Almoosa, N.; Poon, K. Flexible gesture recognition using wearable inertial sensors. In Proceedings of the 2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS), Abu Dhabi, United Arab Emirates, 16–19 October 2016; pp. 1–4.
9. Suri, K.; Gupta, R. Convolutional neural network array for sign language recognition using wearable IMUs. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019; pp. 483–488.
10. Khassanov, Y.; Imanberdiyev, N.; Varol, H.A. Inertial motion capture based reference trajectory generation for a mobile manipulator. In Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, Bielefeld, Germany, 3–6 March 2014; pp. 202–203.
11. Khassanov, Y.; Imanberdiyev, N.; Varol, H.A. Real-time gesture recognition for the high-level teleoperation interface of a mobile manipulator. In Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, Bielefeld, Germany, 3–6 March 2014; pp. 204–205.
12. Digo, E.; Gastaldi, L.; Antonelli, M.; Pastorelli, S.; Cereatti, A.; Caruso, M. Real-time estimation of upper limbs kinematics with IMUs during typical industrial gestures. *Procedia Comput. Sci.* **2022**, *200*, 1041–1047. [CrossRef]
13. Neto, P.; Simão, M.; Mendes, N.; Safeea, M. Gesture-based human-robot interaction for human assistance in manufacturing. *Int. J. Adv. Manuf. Technol.* **2019**, *101*, 119–135. [CrossRef]
14. Kulkarni, P.V.; Illing, B.; Gaspers, B.; Brüggemann, B.; Schulz, D. Mobile manipulator control through gesture recognition using IMUs and Online Lazy Neighborhood Graph search. *ACTA IMEKO* **2019**, *8*, 3–8. [CrossRef]
15. Assad, C.; Wolf, M.T.; Karras, J.; Reid, J.; Stoica, A. JPL BioSleeve for gesture-based control: Technology development and field trials. In Proceedings of the 2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA), Woburn, MA, USA, 11–12 May 2015; pp. 1–6.
16. Wang, W.; Li, R.; Diekel, Z.M.; Chen, Y.; Zhang, Z.; Jia, Y. Controlling object hand-over in human–robot collaboration via natural wearable sensing. *IEEE Trans. Hum.-Mach. Syst.* **2018**, *49*, 59–71. [CrossRef]
17. Hassan, H.F.; Abou-Loukh, S.J.; Ibraheem, I.K. Teleoperated robotic arm movement using electromyography signal with wearable Myo armband. *J. King Saud Univ.-Eng. Sci.* **2020**, *32*, 378–387. [CrossRef]
18. Chico, A.; Cruz, P.J.; Vásconez, J.P.; Benalcázar, M.E.; Álvarez, R.; Barona, L.; Valdivieso, Á.L. Hand Gesture Recognition and Tracking Control for a Virtual UR5 Robot Manipulator. In Proceedings of the 2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM), Cuenca, Ecuador, 12–15 October 2021; pp. 1–6.
19. Fang, B.; Sun, F.; Liu, H.; Guo, D.; Chen, W.; Yao, G. Robotic teleoperation systems using a wearable multimodal fusion device. *Int. J. Adv. Robot. Syst.* **2017**, *14*, 1729881417717057. [CrossRef]
20. Kim, M.; Lee, D. Development of an IMU-based foot-ground contact detection (FGCD) algorithm. *Ergonomics* **2017**, *60*, 384–403. [CrossRef] [PubMed]
21. Knudson, D.V.; Knudson, D. *Fundamentals of Biomechanics*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 183.
22. Alazrai, R.; Mowafi, Y.; Lee, C.G. Anatomical-plane-based representation for human–human interactions analysis. *Pattern Recognit.* **2015**, *48*, 2346–2363. [CrossRef]
23. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
24. Elman, J.L. Finding structure in time. *Cogn. Sci.* **1990**, *14*, 179–211. [CrossRef]
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]