Article

Radio-Frequency-Identification-Based 3D Human Pose Estimation Using Knowledge-Level Technique

Saud Altaf [1,*], **Muhammad Haroon** [1], **Shafiq Ahmad** [2], **Emad Abouel Nasr** [2], **Mazen Zaindin** [3], **Shamsul Huda** [4] **and Zia ur Rehman** [1]

1 University Institute of Information Technology, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi 46300, Pakistan

2 Industrial Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia

3 Department of Statistics and Operations Research, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

4 School of Information Technology, Deakin University, Burwood, VIC 3128, Australia

* Correspondence: saud@uaar.edu.pk; Tel.: +92-300-9466907

Abstract: Human pose recognition is a new field of study that promises to have widespread practical applications. While there have been efforts to improve human position estimation with radio frequency identification (RFID), no major research has addressed the problem of predicting full-body poses. Therefore, a system that can determine the human pose by analyzing the entire human body, from the head to the toes, is required. This paper presents a 3D human pose recognition framework based on ANN for learning error estimation. A workable laboratory-based multisensory testbed has been developed to verify the concept and validation of results. A case study was discussed to determine the conditions under which an acceptable estimation rate can be achieved in pose analysis. Using the Butterworth filtering technique, environmental factors are de-noised to reduce the system's computational cost. The acquired signal is then segmented using an adaptive moving average technique to determine the beginning and ending points of an activity, and significant features are extracted to estimate the activity of each human pose. Experiments demonstrate that RFID transceiver-based solutions can be used effectively to estimate a person's pose in real time using the proposed method.

Keywords: 3D human pose estimation; RFID; filtering; kinematic; ANN

1. Introduction

The human activity recognition (HAR) system has shown tremendous improvement over the past several years in terms of its ability to facilitate communication between humans and machines. The HAR architecture introduces numerous innovations that significantly enhance the ways in which humans and machines communicate with one another. Because of state-of-the-art research and the expansion of a wide variety of input devices to capture data, the process of recognition is becoming less complicated and more useful. Input devices make possible the visualization or detection of human poses in situations ranging from simple to complicated. Radio frequency identification (RFID)-based devices are a good example of this type of technology. These devices are able to accurately identify fine-grained movements despite the presence of complex backgrounds [1]. In order to come up with new ways of making computers more interactive with minimal physical contact, researchers have proposed RFID-based wireless sensing systems [2]. The development of a system that is capable of identifying human poses in a three-dimensional environment through radio frequency technology is one example of this type of progress.

Human pose estimation provides a graphical depiction of a person in a given position. Estimating a person's pose amounts to generating a set of coordinates that may be connected

in various ways to give a full picture of where they are. A skeletal coordinate, or "joint," is any one particular place on the skeleton. A proper connection is a combination of two components of the body that should function together. Unfortunately, not all possible permutations of parts can produce usable pairs.

Advancements in RF sensing systems have generated growing interest in the research and technology application of 3D human pose estimation. RF sensors, as compared with ordinary vision sensors, are unaffected by either light or darkness and have the distinct ability to protect user privacy. Because of their compact design, RFID tags are also considered suitable for deployment as wearable sensors as well as contactless sensing devices. RFID systems are much cheaper than radar-based systems such as the FMCW radar [3].

However, because of the diversity and complexity of wireless channels, it is usually hard to generalize a trained RF sensing system to new environments. RF signals propagate through the air, and the receiver end needs precise signal strength of the deployed location. The received signal strength (RSS) is dependent on many different factors, the most important of which are the placement of the antennas, the surrounding obstacles, the layout of the room walls, and the movement of the object being observed [3]. The focus of the current RFID-based pose detection systems [4,5] is on tracking the movement of a single body part at a time. These systems obtain their phase data from tags that are attached to various parts of the body. If multiple body parts move simultaneously, it can cause inter-tag collisions and an RFID mutual coupling effect, both of which significantly impair the accuracy of the system. For that reason, utilizing RFID tags to track the entire body remains a challenging task. Whenever the surrounding environment varies, identical human subjects performing the same activity can yield RF properties that are significantly different from one another. Even if the same person does the same activity, environmental variations may create differing radio frequency (RF) qualities. Developing 3D human posture estimation algorithms that are environment-aware is a challenge [6].

Many methods have been proposed over the years to improve human–computer interaction (HCI) by researchers. Real-time human pose estimation is useful in a variety of domains, particularly healthcare. A primary motivation in the medical field is to minimize the transmission of environmental contamination by eliminating device contact and monitoring patients' indoor daily activities. Depending on the information obtained from various sensors, these procedures vary. Both fixed and moveable sensors have been widely used for human activity recognition. Both stationary (those permanently affixed to the ground) and mobile (those easily moved from one location to another) sensors are employed to collect information about the study's issue. External sensors can take the form of anything from a camcorder to a mic to a motion sensor to an imaging system to a trigger to an RFID tags chip. Wearable sensors can measure motion and orientation using devices such as gyroscopes, accelerometers, and motion detectors.

Recent research on RFID-based estimation of human pose has revealed the following limitations:

- When the receiver and transmitter are not in close proximity, the observed phase value may not accurately reflect the relationship between path length and received phase [4].
- Recent RFID-based strategies only assess upper-body movement patterns, so it can be difficult to track the entire body at once with them to achieve the required accuracy [5].
- Learning models may not achieve optimal results when applied to a novel RF environment due to the fact that each training variable is based on a relatively small number of datasets. It can be difficult to reconstruct poses from a small dataset when a similar participant is asked to perform the same task in multiple locations, resulting in vastly different RF data [6].
- Before the system can better adapt to different environments, it must overcome the substantial challenge of generalizing the learning model [7].

In order to evaluate the validity of the proposed system, real-time scenarios are considered and compared to existing RFID-Pose systems. According to the findings of the

experiments, the system is capable of accurately tracking three-dimensional human poses for a variety of subjects and shows great subject adaptation.

This study makes several significant contributions, which are briefly summarized below:

- This study presents an environment-adaptive 3D human pose prediction model using transceiver-based RFID tagging on the human body to overcome the problem of only being able to collect a single-phase sample from a single tag.
- A 3D human pose estimation framework is proposed based on the artificial neural network (ANN) as a knowledge-level technique to estimate the learning error.
- A prototype with commercial RFID tags attached to the entire human body has been developed to generate a dataset with ground-truth values for training and evaluating the model.
- An analysis of the variability of RFID data is conducted using the fast Fourier transform (FFT) that has been measured and identifies the primary difficulties associated with generalization issues.
- Case study results indicate that the proposed RFID system can predict 3D human postures with ease and is also highly adaptable. In addition, these results are compared to other published work in the same field to demonstrate their superiority and validate the concept.

In the next sections of this study, Section 2 analyses the published research on the proposed system's development as well as the challenges that it needs to overcome. In Section 3, a mathematical model is used to briefly explain the proposed framework for recognizing human poses. The discussion of the testbed setting and the findings can be found in Section 4. In Section 5, the conclusion and future directions are discussed.

2. Literature Review

Pose estimation has numerous potential applications in diverse fields such as medicine, robotics, computer graphics, and video games. Existing literature on pose recognition can be roughly divided into two categories for convenience: both device-based and device-free pose recognition are possible [2].

Device-based recognition widely uses vision sensors such as cameras [4] and Kinect [5,6] to capture body poses in order to interpret the pose. Among both camera and Kinect technologies, Kinect is covering a large area of research as it provides more options to look at the human poses for improved accuracy and efficiency. On the other hand, device-free human pose received an increased use of the signals generated by commercial hardware devices in order to complete the recognition task. These systems are categorized as: radar-based [8] and received signal strength indicator (RSSI) [9]. In the study, [10] a neural network model was adapted for use in an RFID-based wireless sensor network in an effort to reduce the likelihood of collisions occurring within the network. The ANN adds on layers of feature combinations; the research improves with these added layers. The obtained results demonstrated that the ANN model was the most suitable in terms of prediction reliability. CSI-based [11,12], and a combination of both channel state information (CSI) and RSSI based techniques are used [13]. Other useful techniques for device-free human pose estimation are explored in surveys [2,14]. Research in the field is important because it offers fine-grained signal information at the subcarrier level, which has wide applications in computer vision and human representation for semantic parsing [15].

Recent research has combined Kinect-based data with wireless signal-based data to better recognize human poses [16,17]. Yule Ren et al. [3] presented a 3D human pose tracking system that uses the 2D angle of arrival (AoA) of signals reflected off the human body to estimate a 3D skeleton pose made up of a set of joints. There is only one sensor that can provide 2D AoA to identify moving limbs, so the participant was asked to face the sensor during evaluation. If multiple sensors are deployed at right angles, the user can change orientations. While walking, the system may not work well. The study [1] addressed an RFID-based 3D human pose tracking approach that integrates few-shot fine-

tuning and meta-learning. Larger datasets sampled in new situations are needed to achieve satisfactory fine-tuning performance, which increases training data gathering effort and cost.

In [18], the authors combined computer vision and RFID technology in multi-person scenarios to design a more advanced exercise monitoring system. This allowed them to track more information about the participants' workouts. This design was implemented in a smart exercise equipment application by the study using commercially available Kinect cameras and RFID devices. Using RFID phase data, the authors of [19] presented a real-time 3D pose prediction, subject-adaptive, and tracking system. This system would leverage a unique cycle kinematic network to approximate human postures in real-time. The system was built with commercially available RFID readers and tags, and it was evaluated with an RFID-based comparative methodology.

In the paper [20], the author presented a vision-aided, real-time 3D pose evaluation and tracking system. This system utilized a deep kinematic network to approximate human poses in real-time from RFID phase data. This network was trained with the support of computer vision data as labels gathered by Kinect 2.0. Due to the necessity of the original subject skeleton in the training phase, the proposed methodology is compromised when the subject is tested with an untrained subject or in a distinct standing position/environment. In another study [21], a kinematic network is suggested as a way to train models without having to pair RFID and Kinect data. The subject-adaptive system that came out of it was made by learning how to turn sensors data into a skeletal system for each subject. When tested with a known subject, the efficiency of the model is a little lower than with the classical RFID pose tracking method. Because RFID-based pose estimation relies on RFID tags attached to the human body, it can be classified as a device-based method. The study [22] presents a 3D human pose estimation framework based on a relatively new deep learning model that can encode prior knowledge of the human skeleton into the pose construction procedure to improve the estimated joints' match with the human body's skeletal structure. The system consists of nine diffused antennas and requires the subjects to conduct activities at a fixed point. Therefore, the proposed system is restricted to specific applications and is not suitable for daily use.

According to the available literature, the proposed posture system is the first of its kind to use transceiver-based sensors to estimate three-dimensional human poses covering the whole human body. The proposed system uses RFID and computer vision (CV) to accurately estimate human 3D position across several modalities. The comprehensive review can be found in Table 1, and it draws attention to the potential related research that is concentrated on a variety of factors that affect human pose estimation.

Table 1. Comparison of various related work.

Paper	Estimation System	Hardware	# of RFID Tags	Technique	Tracking Error	Accuracy	Limitations
[1]	Meta Pose	1 reader antennas, 2 RFID Readers	12	Shoulders to knees	5.1 cm	N/A	Phase offset
[3]	RFID Pose	1 reader antennas, 3 RFID Readers	12	Upper body	6.7 cm	95.4%	Adaptability
[19]	Cycle Pose	3 reader antennas, 3 RFID Readers	10	Non head, toe	4.9 cm	N/A	Generalization
[20]	Meta-Leaning	3 antennas, 3 RFID Readers	12	Shoulders to knees	4.5 cm	95.8%	Missing sample Generalization
[21]	Subject adaptive	2 antennas, 3 RFID Readers	12	Upper Body	8.6 cm	N/A	Phase Offset
Our work	Subject and environment adaptive	8 transceivers	8	Whole body (head to toe)	3.46 cm	96.7%	Discussed in future directions section

3. Materials and Methods

This article evaluates a real-world scenario in which a human subject was observed by RFID-based transceivers and a Kinect device in order to construct and analyse the subject's 3D skeleton. The data collected using RFID readers can be used to generate the 3D skeleton of the subject, and the data acquired using the Kinect sensor can be used as ground data for supervised learning. The feed-forward back-propagation neural network is proposed for estimating the human poses. The proposed system consists of three primary components: data collection, data processing, and pose estimation, as shown in Figure 1.
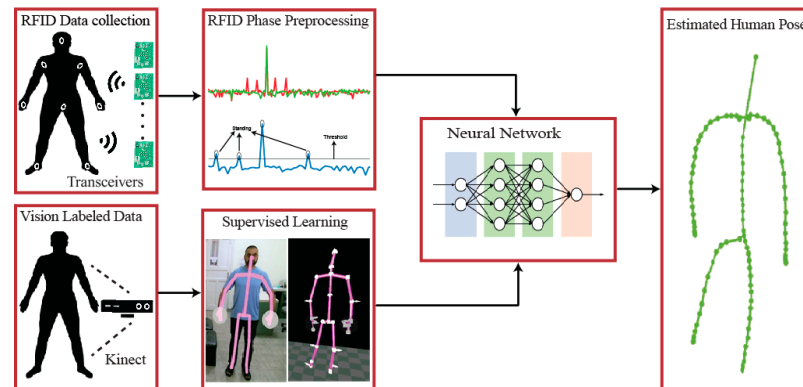
Figure 1. Proposed human pose analysis framework.

3.1. Data Collection

During this phase, data is collected from each of the RFID sensors and then processed in order to construct a three-dimensional skeleton of the subject. The RFID transceivers and the Kinect 2.0 sensors work together to collect the necessary information for testing and training. The data collected from the RFID tags is preprocessed before feature extraction and pose generation. Furthermore, the kinematic information will be used as labelled data for the purpose of conducting supervised training. The RGB camera and the infrared sensors present in the Kinect device conduct an analysis on the three-dimensional position of each human joint, and the findings of this analysis are then saved in a database. For the purpose of the study, passive RFID tags were attached to each of the eight joints of the human body. In order to collect the phase data from all of the linked RFID tags, a total of eight transceivers are utilized as part of the data collection process.

RF sensors were used at a rate of 0–1000 Hz, and each sensor was set to a certain angle on a joint of a human body part between different points of interest. Researchers have collected the samples at frequencies ranging from 5 Hz up to 512 Hz with, essentially, the same test setups [7]. This study proposes a specific frequency range in order to obtain fine-grained human pose movements. For the valid case study, data from an office and laboratory setting were collected to make the proposed system more adaptive to varied environments.

The antenna transmits the RF signal, which is received by the RFID tag and then reflected back to the receiving antenna; this process is described as:

$$r = H\gamma + n \tag{1}$$

where r is the receive vector, H is the channel matrix, γ denotes the backscattering signal at the tag and n is the noise vector.

Figure 2 shows the RFID forward and backward links. The forward link is the transmitter-to-tag transmission channel. A reverse link propagates from the tag to the reader's receiver. Denote the channel gains of the forward and backward links as h^f and h^b, respectively. Then, the whole channel gain can be written as:
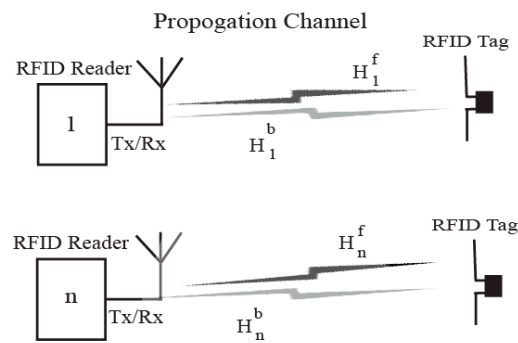
$$H = h^b h^f \gamma \tag{2}$$

Figure 2. An illustration of multiple-tag RFID system.

The relationship between h^f and h^b depends on the transmitter and reader locations. In a monostatic system, transceiver antennas are close together [22]. As forward and backward links are highly correlated, the mutual recognition rule of radio channels suggested in Equation (3) is given by the following:

$$h^b = h^f \tag{3}$$

Let us now look at the channel model of an RFID system with numerous tags and multiple readers. Suppose that N_T tags are attached to the object's body and the reader is equipped with N_{rd} antennas. So, the ith tag is equipped with $N_{tag,i}$ antennas. The channel from the reader to the ith tag and back to the reader again within a time factor can be described using Equation (2), and matrix $H_i(t)$ can be calculated by Equation (4).

$$H_i(t) = h_i^b \, h_i^f \tag{4}$$

$$
\breve{H}_i(t) := \begin{bmatrix}
\left[H_i^f\right]_1 x(t) & 0 & \cdots & 0 \\
0 & \left[H_i^f\right]_2 x(t) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \left[H_i^f\right]_{N_{tag,i}} x(t)
\end{bmatrix} \tag{5}
$$

where h_i^f is the forward channel matrix from the reader to the ith tag, and h_i^b is the backward channel matrix from the ith tag to the reader. Based on Equation (1), the received signal of the reader at time t_k can be written as

$$r(t_k) = \sum_{i=1}^{Nt} H_i(t_k)\gamma_i(t_k) + n(t_k) \tag{6}$$

where,

$$H(t_k) = [H_1(t_k)\, H_2(t_k) \ldots .H_n(t_k), \text{ and } \gamma_i(t_k) = \begin{bmatrix} \gamma_i(t_k) \\ \gamma_i(t_k) \\ \vdots \\ \vdots \\ \gamma_i(t_k) \end{bmatrix}$$

To group all the received signals R of the reader and the transmitted signals S of the tags at different time instants $(t_1, \ldots t_K)$ in Equation (1):

$$R = HS + n \tag{7}$$

where R = [$r(t_1)$, $r(t_2)$, $\cdots r(t_K)$], supposing that the channel does not change within the considered time frame, i.e., H(t_1) = H(t_2) = $\cdots$ = H(t_K) = H, S = [(t_1), (t_2),$\cdots$ (t_K)], and n = [n(t_1), n(t_2),$\cdots$ n(t_K)].

3.2. RFID Data Preprocessing

In order to perform the data preparation, the devices first collect RFID signals, then extract the channel information from those signals, and finally preprocess the data. In particular, the study should begin by de-noising RFID signals in order to remove any noise that may be present. Because the conditions of the channel change, the information regarding the channel requires interpretation on a short-term basis. Where a known signal is transmitted and the channel matrix *H* is estimated, let the training sequence be denoted $P_1, \ldots ., P_N$, where P_i is transmitted over the channel, which can be written as

$$r = H\mathrm{pi} + n \tag{8}$$

To de-noise the acquired signal, this study considered the multipath effect of the RFID signal between a pair of transceivers, which at time *t* and frequency *f* can be expressed as

$$H(f,t) = e^{-j\theta offset} \left[H_s\ (f,t) + \sum_{i\epsilon P_d}^0 \alpha_i(t)e^{-j2\pi f\tau_i(t)}\ \right] \tag{9}$$

where $e^{-j\theta offset}$ is the difference between two waves caused by the carrier frequency difference in receiving and transmitting equipment, $\alpha_i(t)$ is the reduction of the amplitude of a signal, and $\tau_i(t)$ is time of flight for the *i*th path. $H_s(f, t)$ represents the static reflection signals. P_d is the collection of dynamic path components which refer to the signals reflected from moving objects. To remove the noise, the study refers to the method proposed in [11], applying Butterworth filtering between the RFID of multiple antennas:

$$
\begin{aligned}
H_1(f,t)\overline{H}_2(f,t) \\
= H_{1,s}(f,t)\overline{H}_{2,s}(f,t) \\
+ \mathrm{H}_{1,s}\ (\mathrm{f},\mathrm{t}) + \sum_{i\epsilon P_d(2)}^0 \alpha_j(t)e^{-j2\pi f\tau_i(t)} \\
+ H_{2,s}\ (f,t) + \sum_{i\epsilon P_d(1)}^0 \alpha_i(t)e^{-j2\pi f\tau_i(t)} \\
+ \sum_{i\epsilon P_d(1)\ ,\ i\epsilon P_d(2)}^0 \alpha_i(t)\alpha_j(t)e^{-j2\pi f(\tau_i(t)-\tau_j(t))}
\end{aligned}
\tag{10}
$$

3.3. Activity Segmentation

Activity segmentation mainly detects the start and end of an activity and removes the no-activity packets from a sample that corresponds to the whole activity. Since human activity durations are not always the same, this study proposes the adaptive moving average (AMA) filter in order to improve the reliability and accuracy of the real-time pose estimation. The moving average filter allows signals within a selected range of frequencies and time to be processed while preventing unwanted parts of a signal from getting through. The AMA filter averages subsets of the full data set to filter data points. AMA defined for a subset of original signal s(n) is shown in Equation (11).

$$s(n) = \frac{s(n-1) + s(n) + s(n+1)}{3} \tag{11}$$

The adaptive moving average technique works similarly to the sliding window technique in that the entire data set is divided into different segments or windows and the values of each window are compared to the values of the other windows.

Steps to perform the AMA filter are as follows:

- Define sliding window size, shown in Equations (12) and (13).
- Calculate the difference in average ΔA, as shown in Equation (14).

- Calculate the time difference in Δt, as shown in Equation (15).
- Define boundary points array bp[].

To perform the filter, the first step is to set the size of the window, calculated in Equation (12).

$$w = 2f \tag{12}$$

$$bp[j] = i + f, \ \{\forall w[i, i + 2f] \in signal \Delta A \tag{13}$$

where f is the sampling frequency.

The next step is to define the start and end points of a human activity within a signal. First, calculate the difference in averages ΔA between the first half and second half of a sliding window.

$$bp[j] = i + f, \ \left\{ \overline{\Delta A}[i] = \frac{\left| \sum_{i}^{i+f} \Delta A - \sum_{i+f}^{i+2f} \Delta A \right|}{f} \geq th1 \tag{14} \right.$$

where th1 is the threshold point.

Then, calculate the time difference Δt between the two windows from Equation (12), as shown in Equation (14) and Equation (15), respectively.

$$bp[j] = i + f, \ \{t[i] = t[bp[j]] - t[bp[j-1]] > 2s \tag{15}$$

Here, the threshold point is set to 0.5; $i = 1, \ldots n$, n is the length of ΔA signal; $j = 1, \ldots, m$, m is the length of the bp[] array

If the sliding window satisfies both Equations (14) and (15) at the same time, the center point of the window is considered the boundary point stored in the array bp[] to determine the boundary points.

3.4. Channel Feature Selection

The data presented in this article are collected using eight different off-the-shelf transceivers. Let us name the antennas N_t that are used for transmitting the signal at the transmitter's end (T_x), and the antennas N_r that are used for receiving the signal at the receiver's end (R_x). As a result of RFID's use of eight different inputs, the antenna array, which is formed by T_x and R_x components, will generate eight separate data transmission lines.

This study created a feature set containing all the predefined features (M_1, M_2, M_3, ... M_n) extracted from each of the eight received signals about a particular human pose. Predefined features are the values that represent the peaks generated by human activity. For that matter, the amplitude of 0.5 dB is set as the point of threshold. Peaks in the data that are regarded to be the depiction of human pose activities are those that are at or above the threshold point.

The use of amplitude and phase difference as recognition features can better show how body movements affect wireless signals. This is because the amplitude can change, but the phase difference can stay stable for a certain amount of time and can better describe how the frequency of different data streams changes over time. This matrix-based feature set (Fs) contains the number of extracted features, as expressed by Equation (16).

Feature set = [Mean(m), Variance(v), Standard deviation(sd), Average deviation(ad)]

Mean $M_1 = [q_1 \ q_2 \ q_3 \ldots q_n]$

Variance $M_2 = [r_1 \ r_2 \ r_3 \ldots r_n]$

Standard deviation $M_3 = [y_1 \ y_2 \ y_3 \ldots y_n]$

Average deviation $M_4 = [z_1 \ z_2 \ z_3 \ldots z_n]$

$$Feature\ set\ (Fs) = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{bmatrix} \Rightarrow \begin{bmatrix} q_1\ q_2\ q_3 \ldots q_n \\ r_1\ r_2\ r_3 \ldots r_n \\ y_1\ y_2\ y_3 \ldots y_n \\ z_1\ z_2\ z_3 \ldots z_n \end{bmatrix} \Rightarrow \begin{bmatrix} 0\ 1\ 0 \ldots 0 \\ 0\ 0\ 1 \ldots 1 \\ 0\ 0\ 1 \ldots 0 \\ 0\ 1\ 1 \ldots 1 \end{bmatrix} \tag{16}$$

where each entity represents the peak value corresponding to a matrix element 1 and 0. Matrix value 1 indicates the peak value above the threshold point and 0 indicates the peak values below the threshold point.

3.5. Skeleton Construction

This component creates a 3D model of the subject's skeleton using RFID data. Kinematic visual data is used to classify supervised training. The network is trained using a loss function that computes the difference between the estimated posture and labelled vision data, as shown in Equation (17).

$$\epsilon(T) = \frac{1}{8} \sum_{n=1}^{8} \left|\left| \hat{P}_n^T - \dot{P}_n^T \right|\right| \tag{17}$$

where $\hat{P}_n^T$ represents the estimated position, $\dot{P}_n^T$ represents the ground-truth position gathered in the 3D space for joint n at time T, and $\left|\left| \hat{P}_n^T - \dot{P}_n^T \right|\right|$ is the Euclidian distance between these two 3D vectors.

3.6. Classification Phase

Our proposed FFBPN method for training a model, in which the iterations go both ways (feed-forward and back-propagation) to improve the model's performance. Feed forward involves computing input weights in a forward step, and secondly, it adjusts weight and calculates error in a backward step. The data used for training is adjusted to stay between zero and one. The model was trained using 70% of the data set, with the remaining 30% being used for testing and validating the model.

The FFBPN supervised learning begins with an input data matrix Fs denoted by *X*. Each column in *X* represents a single observation. Each column of *X* indicates one predictor or variable. Equation (18) guides model training until the desired predetermined criterion is reached.

$$X_k = \sum_{j}^{n} w_{kj} x_j \tag{18}$$

where X_k represents the updated value of the variable, x_j stands for the previous value, and w_{kj} is the weight link value associated with the neuron/variable. In Equation (19), logsig used as the activation function connecting the input to the hidden layer.

$$f(x) = 1/\left(1 + e^{-x}\right) \tag{19}$$

The positive linear transfer function (POSLIN) used between the hidden layer and the output layer is calculated in Equation (20).

$$f(x) = x \tag{20}$$

Replace missing entries in X with NaN values. The supervised learning methods are capable of handling NaN values, either by ignoring them or by disregarding any row containing a NaN value. The steps for the feed-forward back-propagation network are shown in Algorithm 1.

Algorithm 1: Feed-forward back-propagation network (FFBPN) learning for classification

1:	**Input:** Ds, a dataset containing the training data along with the corresponding targeted values and the learning rate Lr
2:	**Output**: A trained neural network
3:	Initialize all weights and biases in network;
4:	While terminating condition is not satisfied {
5:	for each training tuple X in D_s {
6:	// forward input propagation
7:	for each input layer unit j {
8:	// output of an input unit I its actual input value
9:	$O_j = I_j$;
10:	for each hidden or output layer unit j {
11:	// compute the net input of unit j with respect to the previous Layer, i
12:	$I_j = \sum w_{ij} O_i + \alpha_j$;
13:	// compute the output of each unit j
14:	$O_j = 1/(1 + e^{-x}\, j)$;
15:	// back propagate the errors:
16:	for each unit j in the output layer
17:	// compute the error
18:	$E_j = O_j\,(1 - O_j)\,(T_j - O_j)$;
19:	for each unit j in the hidden layers, from last to first layer
20:	// compute error with respect to the next higher layer, k
21:	$E_j = O_j\,(1 - O_j)\sum E_k\, w_{jk}$
22:	for each weight w_{ij} in network {
23:	// weight increment
24:	$\Delta w_{ij} = (l)\, E_j\, O_i$
25:	// weight update
26:	$w_{ij} = w_{ij} + \Delta w_{ij}$
27:	for each bias α_j in network {
28:	// bias increment
29:	$\Delta \alpha_j = (l)\, E_j$
30:	// bias update
31:	$\alpha_j = \alpha_j + \Delta \alpha_j$
32:	}
33:	}

4. Testbed Environment and Results

Referring to Figure 1, a workable laboratory testbed was developed that consists of eight RF smart sensor modules (XYC-WB-DC transceivers) shown in Figure 3. RF sensors were used at a rate of 1000 Hz, and each sensor was set to a certain angle on a joint of a human body part between different points of interest. Microsoft Kinect 2.0 is used to obtain visual ground truth data for supervised learning and to compare the RF sensors' results. The data was recorded at 30 frames-per-second.
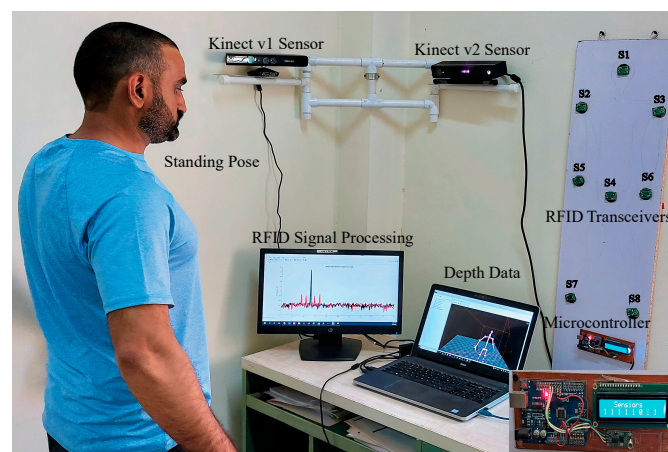

Figure 3. Testbed setup for dataset collection using Kinect and RFID sensors.

For the valid case study, data from an office and laboratory setting were collected to make the proposed system more adaptive to varied environments as shown in Figure 4. There are two indoor environments, office and lab settings, where the distance between the transceivers and the human subjects is between 1 and 2.5 m.
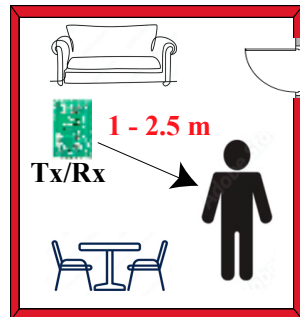

Figure 4. Indoor experimentation setting for human pose acquisition.

As shown in Figure 5a, the points are made up of the head, right shoulder, left shoulder, torso, left hand, right hand, right foot, and left foot. As can be seen in Figure 5b, a total of eight RFID tags were attached to the subject's head, right shoulder, left shoulder, torso, left hand, right hand, right foot, and left foot joints. Even if antennas are used to scan an individual's entire body, all that is necessary for monitoring the majority of human actions is a skeleton with eight joints. RFID tags ALN-9634 that make use of the ultra frequency (UF) are utilized in research by making use of the particular targeted spots of the human body shown in Figure 5b. Using RFID tags and transceivers, the experiments are carried out in a laboratory environment that can be precisely controlled. In order to achieve the highest possible level of efficiency, RFID transceivers incorporate all of the necessary components onto a single circuit board. This enables RFID tag reconfiguration. RFID signals are sensitive to their environment, making it difficult to duplicate and appraise past findings [16]. This research combines RFID signals from four tasks into a dataset (stand, walk, bend, and sit) for the development of a case study. The selected individual performs each task fifty times at a variety of time intervals.

No.	RF Sensor	Name
1	S1	Head
2	S2	Right Shoulder
3	S3	Left Shoulder
4	S4	Torso
5	S5	Right Hand
6	S6	Left Hand
7	S7	Right Foot
8	S8	Left Foot

(**a**)　　　　　　　　　　　　　　　　(**b**)
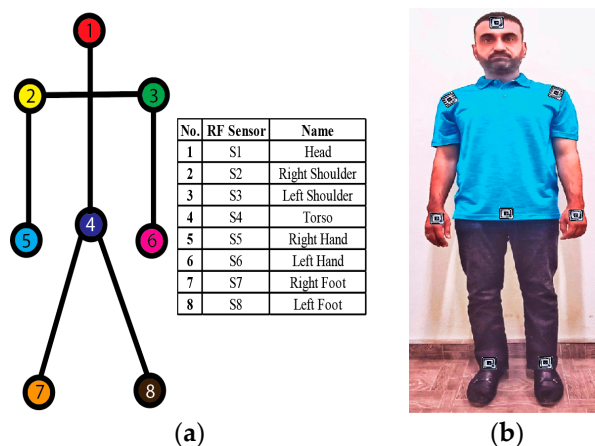
Figure 5. (**a**) Targeted RF points (**b**) RFID tag deployment on subject.

In order to begin the process of data collection, eight RFID transceivers are used to send signals toward respective configured body-connected tags and reflected back to the transceivers. Eight transceivers are used to produce time-domain signals corresponding to a particular human action. Eight signals generated from each transceiver for the walking pose are shown in Figure 6. However, the same number of signals with their corresponding amplitude and frequency are generated for all other human poses discussed in this research.
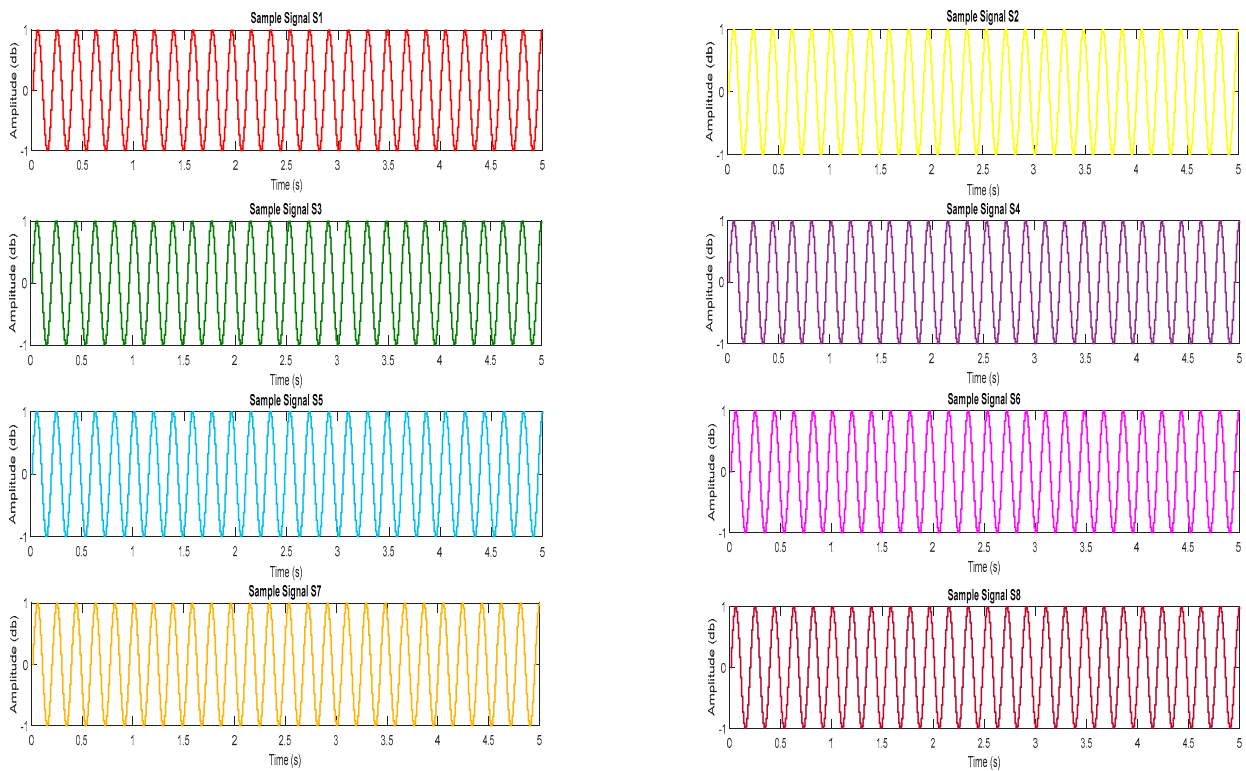
Figure 6. Signals from Eight RFID tags attached on human body performing walking activity.

In order to preprocess these signals efficiently, they are first merged together to form a single signal converted to the frequency domain. Signals are merged together using the Matlab function shown in Figure 7.
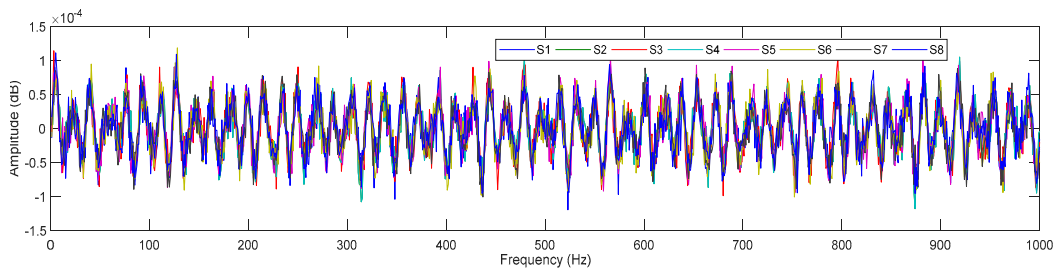

Figure 7. All RFID tags' merged signals.

Figure 7 shows that once the reflected signals from each tag are received at the transceiver, all acquired signals are merged using a Matlab-Merge block script into a signal frequency domain signal.

Here, we assume that there is noise also manifested into the merged signal, and that the original signal may have lost its properties and the system may be confused in further processing. The noise that is created by electrical devices is quite variable, since it is caused by a variety of distinct processes. Figure 5b shows passive tags attached to eight human joints. When interrogating RFID tags, the reader collects phase data using a low-level protocol. To retain the individual identification of all tags, we need to apply an efficient filter to extract the noise from the original signal. Some noises are higher pitched than human poses. To remove out-of-band noise, this study used the Butterworth filter that provides a frequency range in the band-pass filter that will not distort the poses' gesture signals. After that, we use the FFT to illustrate the separation of the noisy signal from the original signal, and then we identify the relevant sideband peaks from the original signal

in order to identify and extract the relevant features. In Figure 8, there are two colors for the signals: red for the signal itself and green for the noise around it.
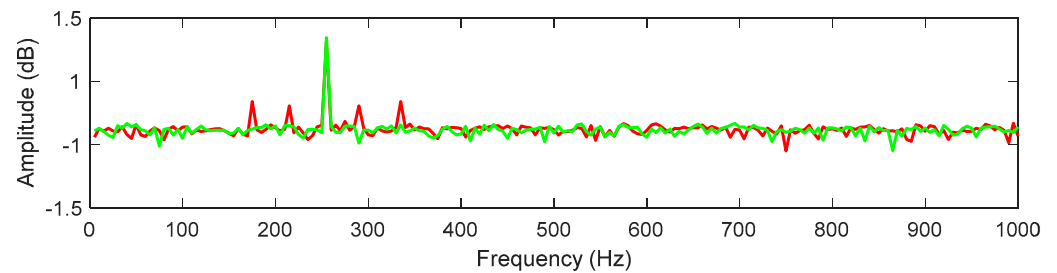
Figure 8. Original signal with noise using FFT.

The multipath effect of RFID signals between two transceivers was treated as noise in this investigation. It is a form of signal reception in which radio signals travel over two or more pathways to reach the antenna. Butterworth filtering was presented as a solution to this problem. The multipath effect was minimized by removing the phase offset data from the merged input signal. Figure 9 shows a filtered image.
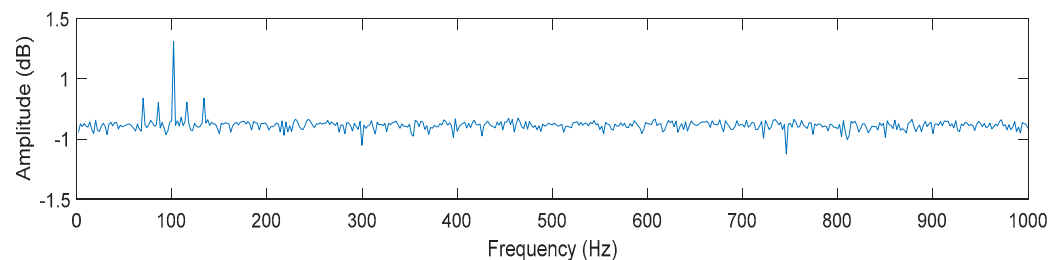
Figure 9. Signal after Butterworth filtering.

The complete activity can be represented by tracing the beginning and ending points of a sample. Our study presents an adaptive moving average (AMA) filter to increase real-time pose estimation, since human activity durations fluctuate. A moving average filter processes signals within a predetermined frequency and time range while excluding unwanted elements of the signal. There is now a clear separation between the segmented signal and other data, as illustrated in Figure 10.
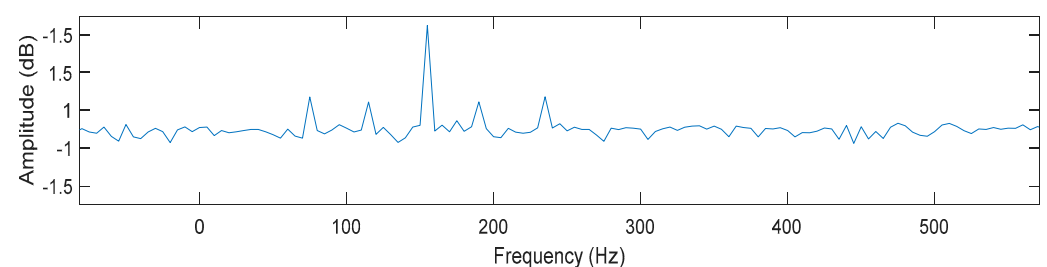
Figure 10. Segmented Signal.

Analyzing the peaks that remain after the segmentation procedure is complete allows for the extraction of characteristics unique to each human activity, as demonstrated in Figure 10. The amplitude of 0.5 dB has been chosen as the point of threshold. Peaks in the data that are regarded to be the depiction of human pose activities are those that are at or above the threshold point. The illustration of the item in its stationary condition is thought to be its peaks when they are lower than the threshold point. A unique signal pattern and a set of peaks are produced as a result of each activity carried out by the object. Both the

number of peaks and the amplitude are determined by the kind of physical activity that is being carried out.

As shown in the following Figure 11a, the walking activity of the item produced at least seven peaks of varying amplitudes over the predetermined threshold. Walking engages more muscle joints than standing, and hence generates more peaks than standing. As illustrated in Figure 11b, the standing position produces four peaks, and the low amplitude peaks are disregarded because they do not correspond to any human stance. Figure 11c depicts the relative characteristics of the object's bending activity as measured by the created peak. The activity of bending caused four peaks to appear that were higher than the predetermined threshold point. As illustrated in Figure 11d, the sitting posture created the fewest number of peaks since it required the least amount of physical movement compared to the other poses studied.
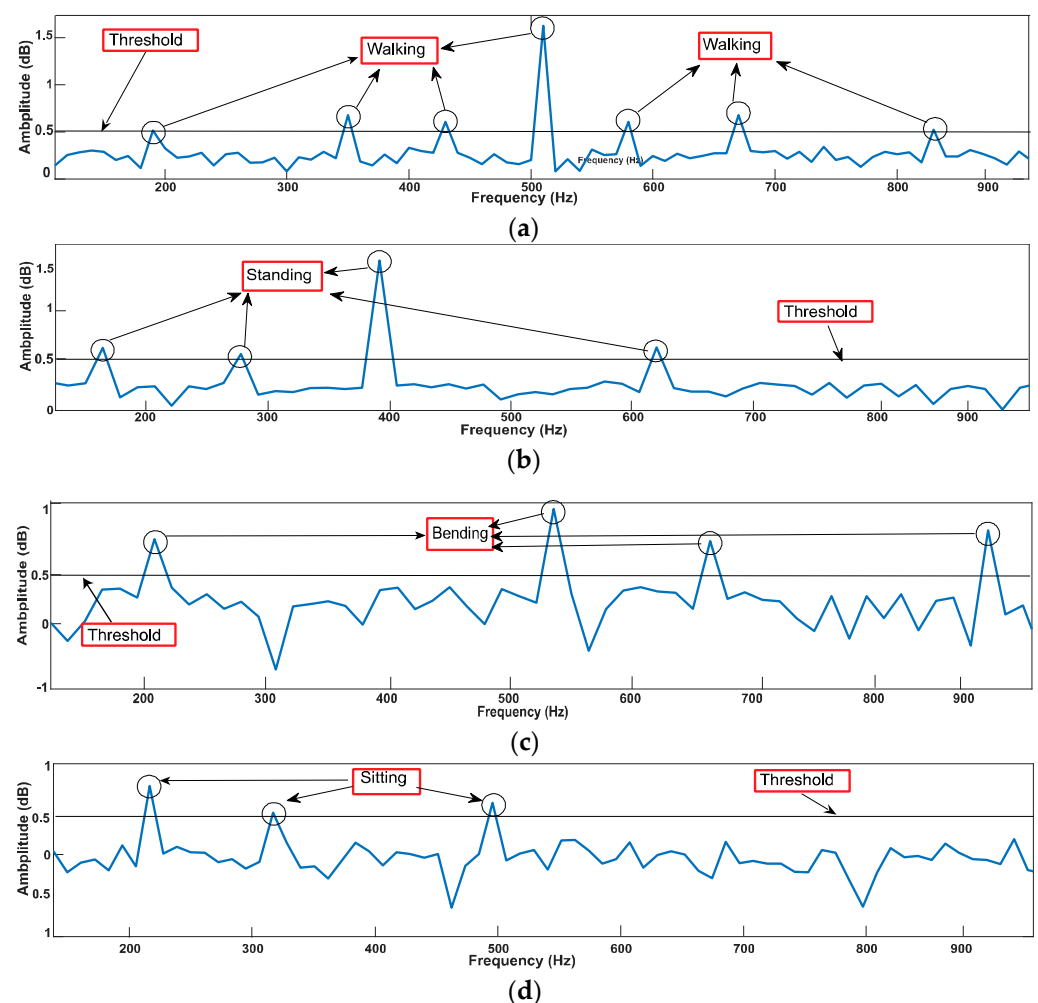
Figure 11. Identification of sideband peaks for pose estimation while (**a**) walking (**b**) standing (**c**) bending (**d**) sitting.

After estimating the pose using RFID signals, they evaluate its precision. The Kinect has the potential to perform 3D bone analysis with significantly greater precision. For the construction of the skeleton, a 320-by-240-pixel image with centimeter-precise depth data is taken and employed. This instrument is totally automated and requires no operator interaction, calibration, or correction. In experiments, a single Kinect camera was positioned around 3 m away from the participant, the minimal distance required to observe the entire human body. Pose data was recorded at 30 Hz. Figure 11 depicts how vision-based data is generated as the ground truth for supervised training.

The kinematic models generated four distinct human activity poses and skeletons, which are depicted in Figure 12a–d. The number of joints formed and their positions change throughout all of the activities. The measurements of human bodies are used in the creation of joints, particularly for the purpose of comparative study. The skeleton that was developed for a body configuration representing walking is depicted in Figure 12a. The skeleton that was obtained for the standing body stance can be seen in Figure 12b. The skeleton that was obtained while the object was in the bending stance is shown in Figure 12c. In addition, Figure 12d illustrates the skeleton that was derived for the seated posture. There is no mechanism for calibrating the Kinect, so the limb lengths are not consistent from frame-to-frame.
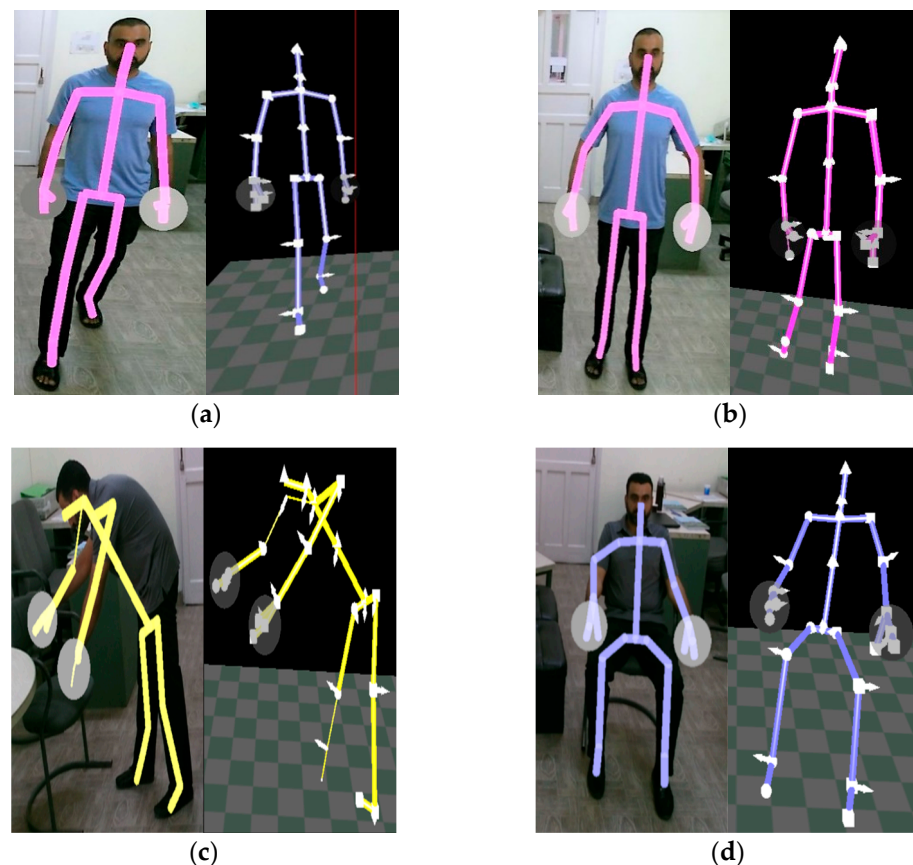
Figure 12. Ground-truth data for (**a**) walking activity (**b**) standing (**c**) bending and (**d**) sitting.

When tested on a specific subject, the advantages of using RFID and vision-based technologies rather than traditional ways to assess human posture become readily apparent. A comparison of the two approaches is shown in Figures 13–16, which depicts the situation in which an untrained individual is executing four predetermined pose activities (that is, walking, standing, bending, and sitting, respectively).
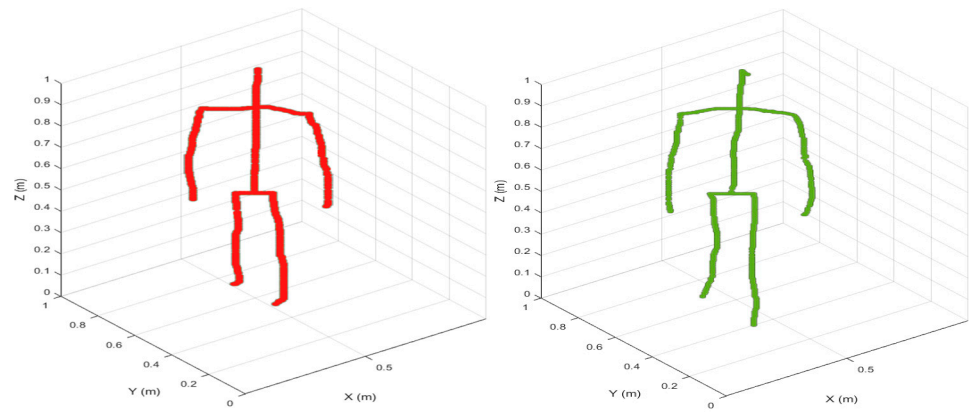
Figure 13. Pose estimation of walking position.
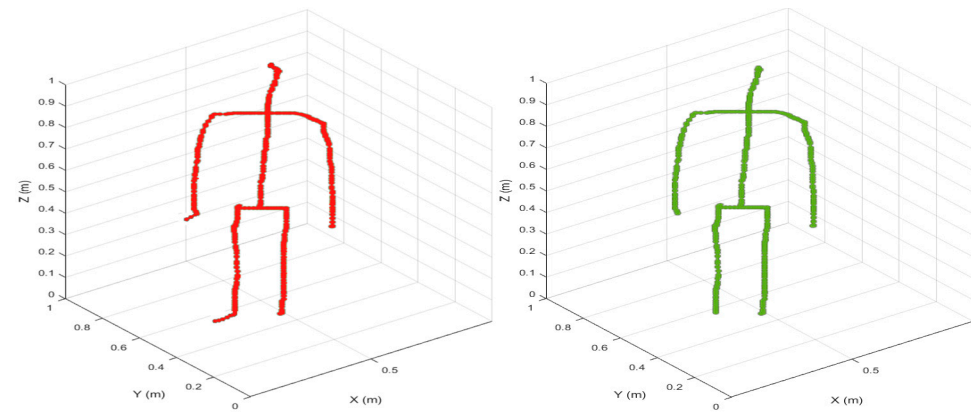
Figure 14. Pose estimation of standing position.
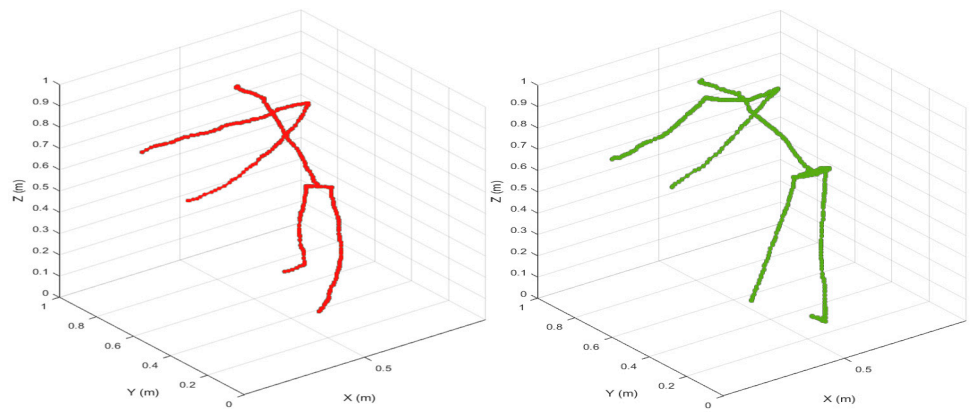

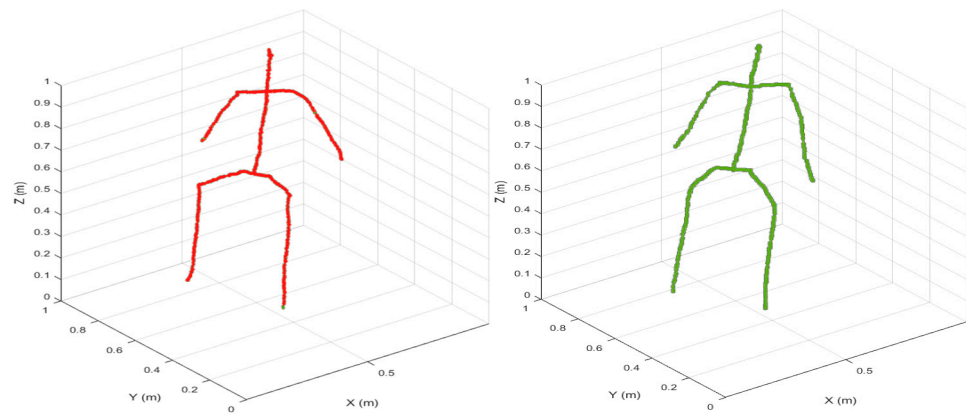**Figure 15.** Pose estimation of bending position.

Figure 16. Pose estimation of sitting position.

In Figures 13–16, the skeleton in red represents an entity that was generated in 3D format using RFID data, and the skeleton in green represents data from a Kinect sensor to determine the error estimation differences between the two types of data. Both of these images illustrate that the skeletons that were reconstructed using RFID and vision-based approaches were extremely comparable to the corresponding ground-truth data. The training data includes validation on four activities that correspond to the following different human poses: walking, standing, bending, and sitting. As seen in Figure 17, the green circles are the reconstructed RFID data, whereas the red dots are the supervised training data.
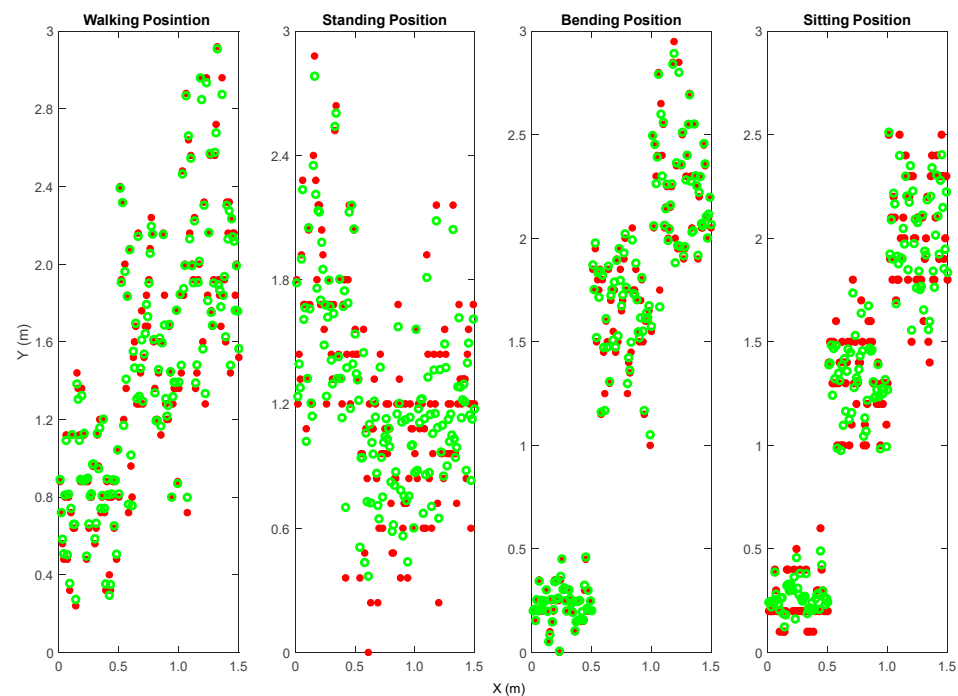
Figure 17. Comparison of reconstructed RFID data and supervised training data.

Illustration of the estimation inaccuracy for different body positions, including walking, standing, bending, and sitting, are shown in Figure 18. The performance was judged based on the nature of this inaccuracy. The precision of the pose estimate depends on the motion being tracked, as indicated in the figure. The biggest inaccuracy occurred when tracking walking action (3.46 cm), while the lowest error was encountered when analyzing sitting position (3.00 cm). The fact that the model has various issues with the joints in its torso is the key factor that contributes to these defects in the model. However, RFID-based

Pose is still accurate for all activities, and the biggest error throughout all tests is relatively smaller than the biggest error that the emerging RFID pose approximation technique can produce. This demonstrates that RFID-Pose is an advancement above the technique that was used previously, i.e., 4.55 cm [20]. The estimation demonstrates that the new RFID-based Pose system can more accurately forecast joint angles and reconstitute the whole body's pose in motion by using RFID phase data. This is reflected in the fact that the device is capable of carrying out the task in question without any problems. The RFID-Pose system had fewer estimation errors than the old method for most motions during validation. This was the situation with every single one of the moves, with the exception of one.
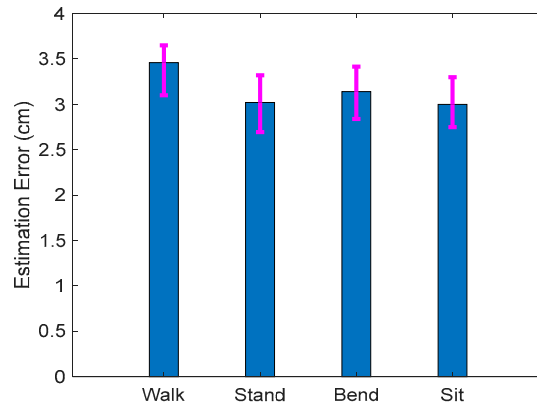
Figure 18. Error estimation for different human poses.

Referring to Equation (16), we calculated the features corresponding to each human pose. The features are mean (m), variance (v), standard deviation (sd), and average deviation (ad) calculation based on peak values as shown in Table 2. The features data are than fed into our neural network as an input, as shown in Figure 19.

Table 2. Feature values calculation.

| Features		Walking	Standing	Bending	Sitting
M_1	m	326.35	313.90	256.15	211.70
	v	16544	12851	13789	18356
	sd	114.79	185.72	193.35	146.17
	ad	106.42	74.89	99.60	103.07
M_2	m	362.96	317.5	332.49	303.35
	v	16387	14658	10525	13107
	sd	110.18	81.52	104.78	76.63
	ad	112.61	80.53	96.15	73.10
M_3	m	434.90	370.25	435.07	409.23
	v	5625	9320	8287	7440
	sd	73.00	89.46	85.96	74.12
	ad	69.78	81.66	70.85	61.80
M_4	m	472.01	505.33	507.75	520.77
	v	5157	4899	3547	6015
	sd	75.03	71.54	61.29	79.71
	ad	72.92	65.56	57.88	71.11
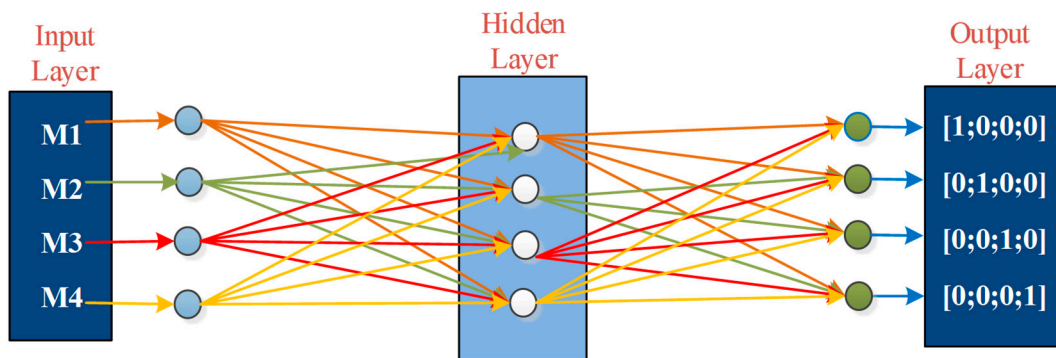
Figure 19. Architecture of proposed artificial neural network-based pose classification.

To validate our model, a number of the hidden layer neurons are selected for the desired result at the output layer. For human pose estimation, the defined output vector classes are written as follows:

- [1; 0; 0; 0]: Human Natural Activity;
- [0; 1; 0; 0]: Human Pose Activity;
- [0; 0; 1; 0]: Unknown Human Pose;
- [0; 0; 0; 1]: No Activity.

A multi-layer feed-forward neural network (FFNN) method is used in this paper for the estimation of human poses. The proposed architecture of ANN for a single hand gesture is presented in Figure 20. Whereas, Table 3 shows the brief explanation and ANN layer setup information.
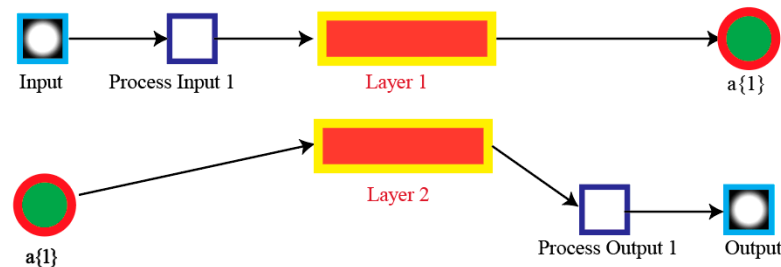
Figure 20. The internal architecture of feed-forward neural network.

Table 3. Description of the implemented ANN.

| NN Steps	Artificial Neural Network Structure for Performance Matrices
Network Model	Feedforward neural network
Training Pattern	Back propagation
Learning Goal	0.001
Input data	Four 1D matrix arrays with data of each class are presented for human pose estimation
Hidden layer neurons	Multiple architectures with different neuron values inside the hidden layer. $[4 \times 10 \times 3]$, $[4 \times 20 \times 3]$, and $[4 \times 30 \times 3]$ (Figure 21).
Target outputs	Mathematical matrices refer to the classified vector classes with value 0 or 1.
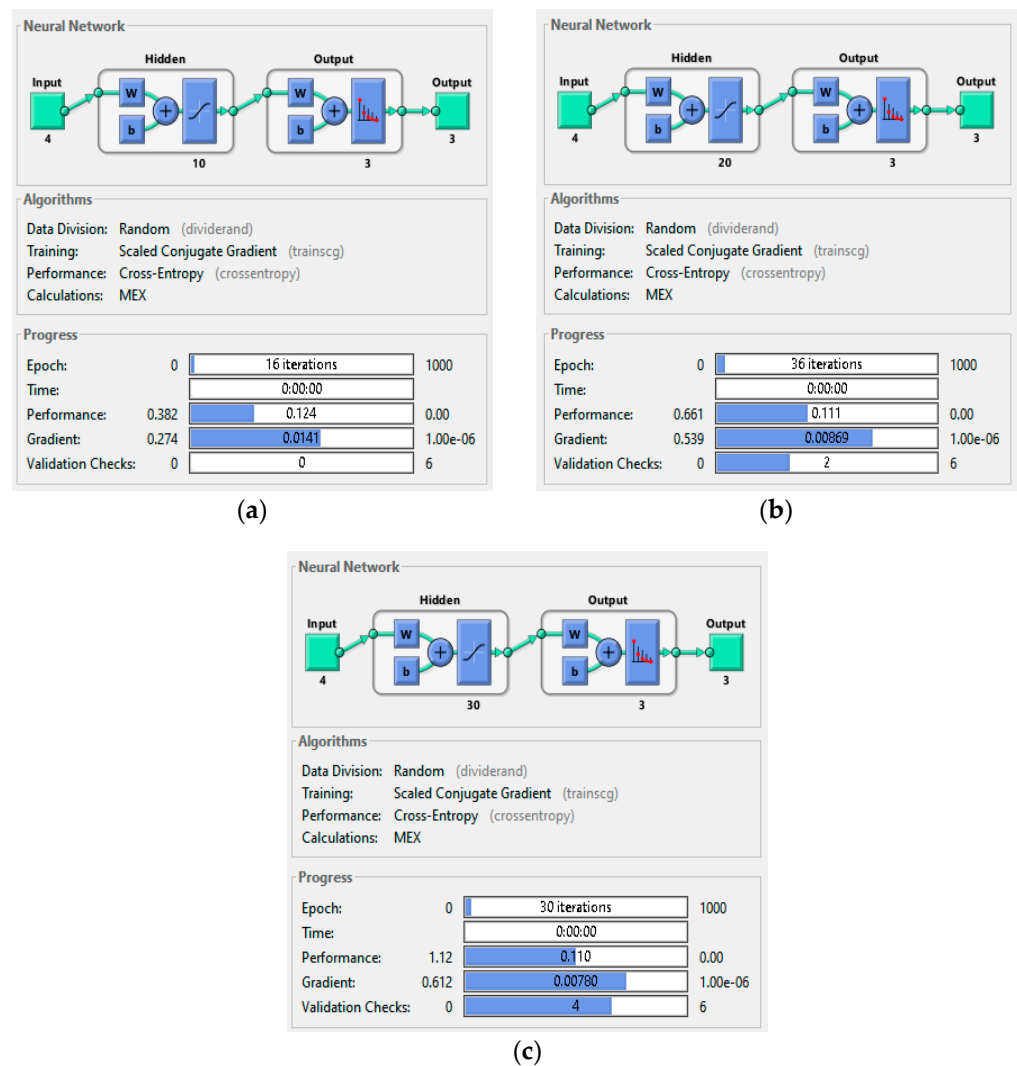
Figure 21. Overview of the different ANN architectures: (**a**) [4 ×10 × 3]; (**b**) [4 × 20 × 3]; (**c**) [4 × 30 × 3].

In this research, three distinct ANN architectures ([4 × 10 × 3], [4 × 20 × 3], and [4 × 30 × 3]) were tested as training tools with the aim of selecting the most suitable hidden layer neuronsas shown in Figure 21. In order to attain the desired result at an acceptable error rate, the hidden layer weights were adjusted until the end result was reached at a reasonable epoch number, as shown in Table 4.

Table 4. Different ANN architecture for classification performance.

Arch	Sample	MSE	No. of Epoch	Accuracy	Classification Error
	M_1	7.34×10^{-2}	80	91.4	8.6
[4 × 10 × 3]	M_2	7.03×10^{-2}	72	93.9	6.1
	M_3	6.37×10^{-2}	70	92.5	7.5
	M_4	7.69×10^{-2}	99	92.7	7.3
	M_1	8.74×10^{-2}	125	97.9	2.1
[4 × 20 × 3]	M_2	8.53×10^{-2}	131	96.4	3.6
	M_3	7.49×10^{-2}	139	96.9	3.1
	M_4	9.06×10^{-2}	147	97.4	2.6
	M_1	7.43×10^{-2}	250	94.5	5.5
[4 × 30 × 3]	M_2	6.00×10^{-2}	284	93.7	6.3
	M_3	7.70×10^{-2}	301	92.8	7.2
	M_4	7.41×10^{-2}	325	92.4	7.6

Table 4 demonstrates that, when compared to alternative ANN architectures, the one chosen [4 × 20 × 3] has higher mean squared error (MSE) efficiency at a suitable number of epochs and error rate.

After the selection of suitable architecture, we can calculate the accuracy using the confusion matrix (CM). Adjusting the hidden layer in the ANN architecture allows the features' input values to be incorporated into the CM's construction. Figure 22 depicts the confusion matrices for walking, sitting, standing, and bending activities.

Figure 22. *Cont.*
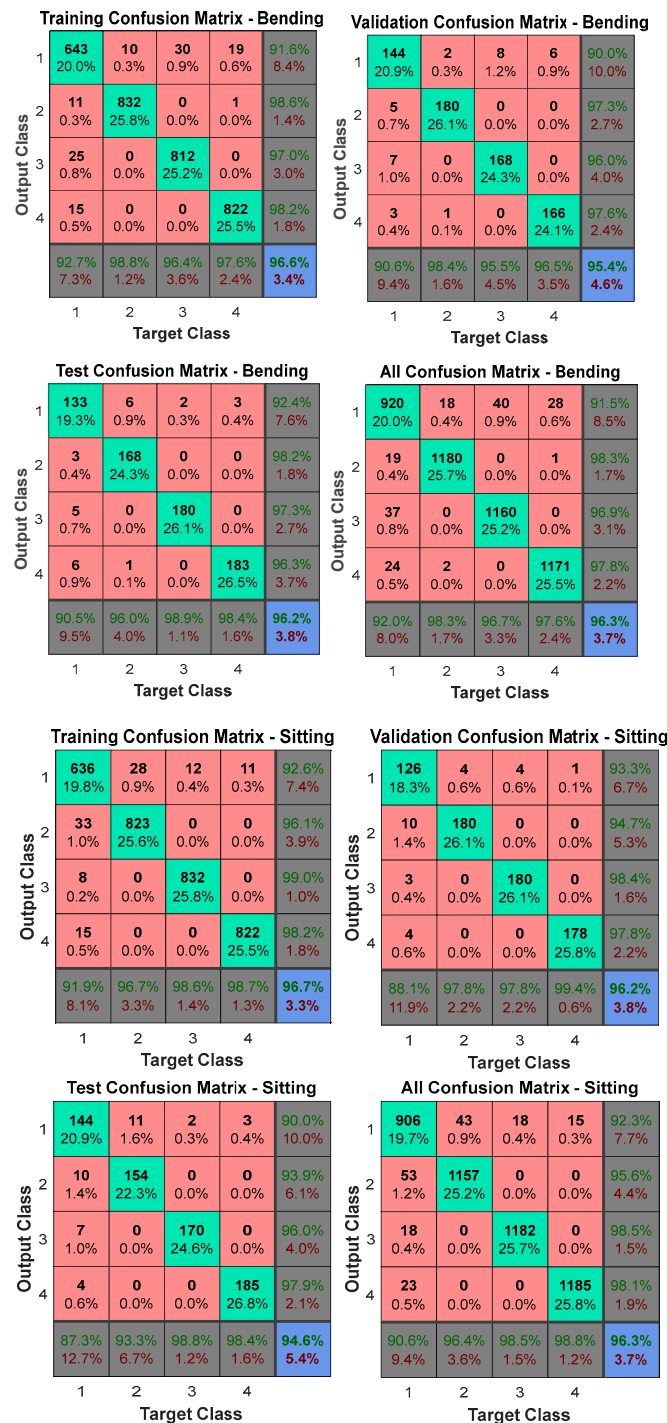
Figure 22. Confusion matrices for walking, standing, sitting, and bending activities.

Each corner cell in the preceding figure depicts a pattern case of an activity that was successfully tested through the proposed ANN architecture used to determine the estimation of human poses. The confusion grid in the confusion matrices graph stores the features-processed training data between the target and output classes, with each of the three phases (preparation, testing, and training) of human pose estimation and individual performance measurement of the ANN architecture comprising its own confusion matrix. From Figure 22, we can see that the walking activity achieved a maximum accuracy rate of 97.8% with only a 2.2% error rate, demonstrating the processing time efficiency of the ANN architecture. Whereas, standing activity achieved a maximum accuracy rate of 97.2% with only a 2.8% error rate, bending activity achieved a maximum accuracy rate of 96.3%

with only a 3.7% error rate, and sitting activity achieved a maximum accuracy rate of 96.3% with only a 3.7% error rate.

Figure 23 shows the overall confusion matrix across all activities. To illustrate the thoroughness of the testing procedure for data validity, four target and vertical output classes were defined to cover the variety of attainable values for the sampled features. Groups of data that have been correctly classified after going through the CM grid's training process are represented by green cells. Each horizontal grey corner cell represents a set of training data that has been successfully tested for its ability to be classified into one of several predefined classes. The red cell displays the data sets that have been incorrectly classified or may not have been adequately validated during the testing phase.

Figure 23. Confusion matrix for identified human activities.

As a final measure, the blue cell displays the sum of all test cases from activities that were correctly classified. Confusion matrices diagrams make it clear that all classes were tested on at least 1200 test instances, with error rates of less than 1% across all trained datasets, as indicated by the percentages displayed in the green cells. Overall, the blue cell achieved a maximum accuracy rate of 96.7% with only a 3.3% error rate, demonstrating the processing time efficiency of the ANN architecture.

Comparison with Baseline Scheme

Finally, a cutting-edge baseline method, namely a meta-learning-based RFID pose tracking system [20], was used to conduct a comparative study. Our research uses the laboratory-collected training and testing dataset. Figure 24 shows the estimation error for each of the different poses. The graph verifies that the performance of both systems is comparable. However, the Meta-learning Pose, when applied to the three unknown poses (i.e., standing, sitting, and bending), generates relatively larger errors. These findings show that the proposed estimation method identifies more accurate initial estimation variables for the new data domains than Meta-Pose.

Figure 24 demonstrates, in addition, that RFID-based pose estimation was able to obtain a greater level of precision while tracking the whole human body than the conventional methods. This is because, when testing different people, the RFID-Pose system works better when cross-skeleton training is used. However, sometimes, traditional joint estimation methods compromise pose recognition accuracy when used to identify skeleton foot position.

The mean estimation error in each untrained data domain is shown in Table 5. The table shows that Meta-Pose has an average error of 4.28 cm across all of the new data domains, while the proposed RFID-based method has an average error of 3.19 cm. In addition, we find that the Meta-Pose estimation error for untrained data domains is still larger.
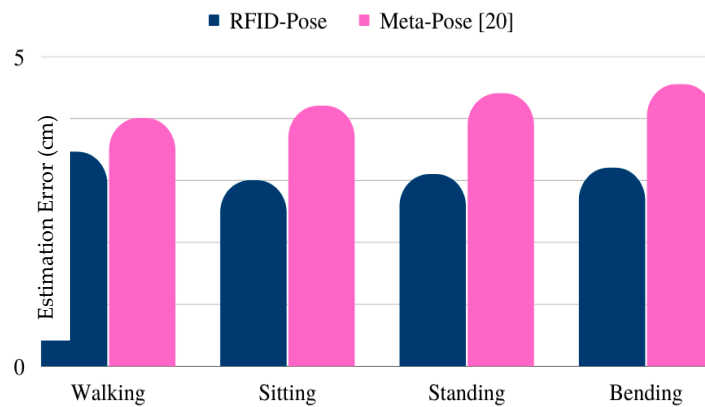
Figure 24. Error estimation comparison for different body poses positions.

Table 5. Performance comparison with mean estimation error.

Poses	RFID-Pose [3]	Cycle-Pose [19]	Meta-Pose [20]	Our Proposed RFID System
Walking	6.72 cm	4.12 cm	4.0 cm	3.46 cm
Sitting	7.62 cm	4.43 cm	4.2 cm	3.0 cm
Standing	5.46 cm	4.51 cm	4.4 cm	3.1 cm
Bending	4.62 cm	4.97 cm	4.55 cm	3.2 cm
Mean Error	6.27 cm	4.50 cm	4.28 cm	3.19 cm

The estimation error for each tagged joint is shown in Figure 25. The joints were numbered from 1 to 8 in the following order: head, right shoulder, left shoulder, torso, left hand, right hand, right foot, and left foot joints. The left and right foot estimation errors were over 3.9 cm for both approaches. This significantly higher set of errors can be attributed, in large part, to the kinematic technique as well as the positioning of the sensors. When computing the location of a joint based on the position of its parent joint, the mistakes from the previous joints will accrue. Because of this, the estimation error of the torso will affect the accuracy of both feet.
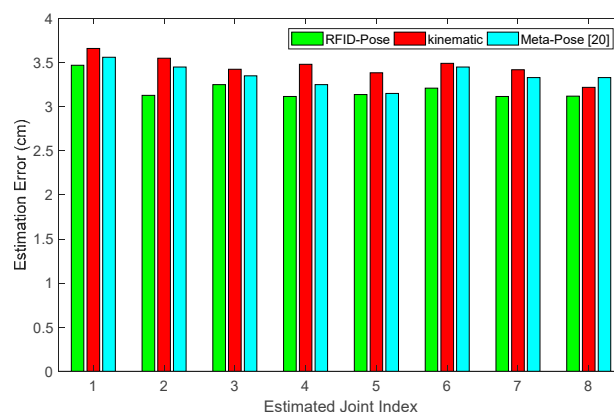

Figure 25. Estimation error for each human joint, numbered from 1 to 8 in the following order: head, right shoulder, left shoulder, torso, right hand, left hand, right foot, and left foot joints.

5. Conclusions and Future Directions

This paper presented an environment-adaptive 3D human pose estimation method employing transceiver-based RFID tagging on the human body. This study conducts an analysis of the variability of the measured RFID data and identifies the primary difficulties associated with generalization issues. At the preprocessing stage, Butterworth filtering is

used to reduce the computational cost by de-noising environmental factors, and adaptive moving average segmentation is used to determine the start and end of an activity. For the valid case study, two kinds of data were gathered from this setup. RF sensors were used at a rate of 1000 Hz, and each sensor was set to a certain angle on a joint of a human body part between different points of interest. Microsoft Kinect 2.0 is used to obtain visual ground truth data for supervised learning and to compare the RF sensors' final results. The data was recorded at 30 frames per second. Data is collected from each of the RFID transceivers and then processed in order to construct a three-dimensional skeleton of the subject. The RFID transceivers and the Kinect 2.0 sensors work together to collect the necessary information for testing and training. The data collected from the RFID tags are preprocessed before feature extraction and pose generation. Furthermore, the kinematic information will be used as labelled data for the purpose of conducting supervised training. The RGB camera and the infrared sensors present in the Kinect device conduct an analysis on the three-dimensional position of each human joint, and the findings of this analysis are then saved in a database. A 3D human pose estimation model is proposed based on artificial neural network (ANN) learning error estimation. The results of a case study demonstrate that the proposed RFID system is able to predict 3D human postures with ease and is extremely adaptable. This research combines RFID signals from four tasks into a dataset (stand, walk, bend, and sit) for the development of a case study. The selected individual performs each task fifty times at a variety of time intervals. After that, we use the FFT to illustrate the separation of the noisy signal from the original signal, and then we identify the relevant sideband peaks from the original signal in order to identify and extract the relevant features. The results demonstrate the estimation inaccuracy for different body positions, including walking, standing, bending, and sitting. The performance was evaluated according to the nature of this error. The precision of the estimated pose is dependent on the tracked motion, as indicated by the provided results. The maximum error (3.46 cm) was encountered when analyzing walking action, while the smallest error was met when analyzing sitting posture (3.00 cm). As shown by the results, the proposed model addresses a variety of issues, including those pertaining to its joints and torso, which is the most important contribution made by other authors. Still, the RFID-based Pose is reliable across the board, and even the largest error observed across all tests is smaller than what can be achieved with the current state-of-the-art in RFID pose approximation. This demonstrates that RFID-Pose is an advancement above the technique that was used previously, i.e., 4.55 cm [20]. The estimation demonstrates that the new RFID-based Pose system can more accurately forecast joint angles and reconstitutes the whole body's pose in motion by using RFID phase data. Furthermore, these results are compared with other related published work to show better efficiency and prove the concept.

In terms of future 3D human pose estimation improvements, the following additional advancements could be researched to improve overall system operation:

- This study analyses RFID data variability and generalization concerns. The generalization issue could be reduced by expanding the training dataset to include additional subjects and positions. Future study will continue to address the generality challenges of RFID-based pose monitoring systems.
- It is important to sample a larger dataset of multiple objects with diverse poses in the different environments in order to obtain a level of performance in fine-tuning that is considered to be satisfactory.
- The 3D human posture estimation system that is built on a cloud-edge framework could potentially be enhanced with the addition of hybrid artificial intelligence approaches.
- Multiple human objects must be considered concurrently with additional poses and machine learning techniques.

References

1. Yang, C.; Wang, L.; Wang, X.; Mao, S. Meta-Pose: Environment-adaptive Human Skeleton Tracking with RFID. In Proceedings of the IEEE GLOBECOM 2022, Rio de Janeiro, Brazil, 4–8 December 2022; pp. 1–6. [CrossRef]
2. Liu, J.; Teng, G.; Hong, F. Human Activity Sensing with Wireless Signals: A Survey. *Sensors* **2020**, *20*, 1210. [CrossRef] [PubMed]
3. Yang, C.; Wang, X.; Mao, S. RFID-Pose: Vision-Aided Three-Dimensional Human Pose Estimation With Radio-Frequency Identification. *IEEE Trans. Reliab.* **2021**, *70*, 1218–1231. [CrossRef]
4. Badiola-Bengoa, A.; Mendez-Zorrilla, A. A Systematic Review of the Application of Camera-Based Human Pose Estimation in the Field of Sport and Physical Exercise. *Sensors* **2021**, *21*, 5996. [CrossRef] [PubMed]
5. Lin, K.-C.; Ko, C.-W.; Hung, H.-C.; Chen, N.-S. The effect of real-time pose recognition on badminton learning performance. *Interact. Learn. Environ.* **2021**, 1–15. [CrossRef]
6. Haroon, M.; Altaf, S.; Ahmad, S.; Zaindin, M.; Huda, S.; Iqbal, S. Hand Gesture Recognition with Symmetric Pattern under Diverse Illuminated Conditions Using Artificial Neural Network. *Symmetry* **2022**, *14*, 2045. [CrossRef]
7. Khusainov, R.; Azzi, D.; Achumba, I.E.; Bersch, S.D. Real-Time Human Ambulation, Activity, and Physiological Monitoring: Taxonomy of Issues, Techniques, Applications, Challenges and Limitations. *Sensors* **2013**, *13*, 12852–12902. [CrossRef] [PubMed]
8. Ding, W.; Guo, X.; Wang, G. Radar-Based Human Activity Recognition Using Hybrid Neural Network Model With Multidomain Fusion. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *57*, 2889–2898. [CrossRef]
9. Oguchi, K.; Maruta, S.; Hanawa, D. Human Positioning Estimation Method Using Received Signal Strength Indicator (RSSI) in a Wireless Sensor Network. *Procedia Comput. Sci.* **2014**, *34*, 126–132. [CrossRef]
10. Mafamane, R.; Ouadou, M.; Sahbani, H.; Ibadah, N.; Minaoui, K. DMLAR: Distributed Machine Learning-Based Anti-Collision Algorithm for RFID Readers in the Internet of Things. *Computers* **2022**, *11*, 107. [CrossRef]
11. Wang, Y.; Guo, L.; Lu, Z.; Wen, X.; Zhou, S.; Meng, W. From Point to Space: 3D Moving Human Pose Estimation Using Commodity WiFi. *IEEE Commun. Lett.* **2021**, *25*, 2235–2239. [CrossRef]
12. Kato, S.; Fukushima, T.; Murakami, T.; Abeysekera, H.; Iwasaki, Y.; Fujihashi, T.; Watanabe, T.; Saruwatari, S. CSI2Image: Image Reconstruction From Channel State Information Using Generative Adversarial Networks. *IEEE Access* **2021**, *9*, 47154–47168. [CrossRef]
13. Yan, J.; Ma, C.; Kang, B.; Wu, X.; Liu, H. Extreme Learning Machine and AdaBoost-Based Localization Using CSI and RSSI. *IEEE Commun. Lett.* **2021**, *25*, 1906–1910. [CrossRef]
14. Wu, D.; Zhang, D.; Xu, C.; Wang, H.; Li, X. Device-Free WiFi Human Sensing: From Pattern-Based to Model-Based Approaches. *IEEE Commun. Mag.* **2017**, *55*, 91–97. [CrossRef]
15. Zhou, T.; Wang, W.; Liu, S.; Yang, Y.; Van Gool, L. Differentiable Multi-Granularity Human Representation Learning for Instance-Aware Human Semantic Parsing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 1622–1631. [CrossRef]
16. Guo, L.; Wang, L.; Liu, J.; Zhou, W.; Lu, B. HuAc: Human Activity Recognition Using Crowdsourced WiFi Signals and Skeleton Data. *Hindawi J. Wirel. Commun. Mob. Comput.* **2018**, *2018*, 6163475. [CrossRef]
17. Ren, Y.; Wang, Z.; Tan, S.; Chen, Y.; Yang, J. Winect: 3D human pose tracking for free-form activity using commodity WiFi. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–29. [CrossRef]

18. Liu, Z.; Liu, X.; Li, K. Deeper Exercise Monitoring for Smart Gym using Fused RFID and CV Data. In Proceedings of the IEEE INFOCOM 2020, Toronto, ON, Canada, 6–9 July 2020; pp. 11–19. [CrossRef]
19. Yang, C.; Wang, X.; Mao, S. RFID-based 3D human pose tracking: A subject generalization approach. *Digit. Commun. Netw.* **2021**, *8*, 278–288. [CrossRef]
20. Yang, C.; Wang, L.; Wang, X.; Mao, S. Environment Adaptive RFID-Based 3D Human Pose Tracking With a Meta-Learning Approach. *IEEE J. Radio Freq. Identif.* **2022**, *6*, 413–425. [CrossRef]
21. Yang, C.; Wang, X.; Mao, S. Subject-adaptive Skeleton Tracking with RFID. In Proceedings of the 2020 16th International Conference on Mobility, Sensing and Networking, MSN 2020, Tokyo, Japan, 17–19 December 2020; pp. 599–606. [CrossRef]
22. Zheng, F.; Kaiser, T. *Digital Signal Processing for RFID*; Wiley: New York, NY, USA, 2016. [CrossRef]