Digital Communications and Networks 8 (2022) 278–288

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.keaipublishing.com/dcan

RFID-based 3D human pose tracking: A subject generalization approach

Chao Yang [a], Xuyu Wang [b], Shiwen Mao [a,*]

[a] Dept. of Electrical and Computer Engineering, Auburn University, Auburn, AL, 36849-5201, USA
[b] Dept. of Computer Science, California State University, Sacramento, CA, 95819-6021, USA

ARTICLE INFO

Keywords:
Radio-frequency identification (RFID)
Three-dimensional (3D) human pose tracking
Cycle-consistent adversarial network
Generalization

ABSTRACT

Three-dimensional (3D) human pose tracking has recently attracted more and more attention in the computer vision field. Real-time pose tracking is highly useful in various domains such as video surveillance, somatosensory games, and human-computer interaction. However, vision-based pose tracking techniques usually raise privacy concerns, making human pose tracking without vision data usage an important problem. Thus, we propose using Radio Frequency Identification (RFID) as a pose tracking technique via a low-cost wearable sensing device. Although our prior work illustrated how deep learning could transfer RFID data into real-time human poses, generalization for different subjects remains challenging. This paper proposes a subject-adaptive technique to address this generalization problem. In the proposed system, termed Cycle-Pose, we leverage a cross-skeleton learning structure to improve the adaptability of the deep learning model to different human skeletons. Moreover, our novel cycle kinematic network is proposed for unpaired RFID and labeled pose data from different subjects. The Cycle-Pose system is implemented and evaluated by comparing its prototype with a traditional RFID pose tracking system. The experimental results demonstrate that Cycle-Pose can achieve lower estimation error and better subject generalization than the traditional system.

1. Introduction

With the rapid development of computer vision, human pose tracking has become an important problem area in recent years, evolving from two-dimensional (2D) poses to three-dimensional (3D) poses [1,2]. Although camera-based techniques have been demonstrated as effective, such vision-based techniques frequently raise security and privacy concerns. For example, the millions of wireless security cameras deployed worldwide have been reported as susceptible to hacking [3], while video data used for pose tracking can also be intercepted and used illegally. To address this concern, Radio Frequency (RF) sensing-based schemes have been proposed using various wireless communication technologies, such as WiFi [4,5], Frequency-Modulated Continuous Wave (FMCW) radar [6], and mmWave radar [7].

To this end, Radio Frequency Identification (RFID) provides a promising RF sensing-based solution for human pose estimation [8,9]. Compared with existing contact-free RF sensing systems, RFID tags can be used as wearable sensors owing to their small form factor. Moreover, RFID systems also have the advantage of being less susceptible to the multipath effect. Compared with advanced radar-based systems, RFID systems are cheaper. However, because RFID data has a lower data/sampling rate than other RF sensing systems, generating a joint confidence map for all joints is significantly more challenging. Therefore, most existing RFID-based pose tracking systems focus on monitoring the movement of one particular limb using sampled phase data [10,11]. In cases when several joints move concurrently, pose detection performance may degrade due to interference from other RFID tags.

Subject adaptability is another challenge in applying RF technology to 3D human pose tracking. Different people have different skeleton forms; however, most neural networks utilized in RF-based pose tracking systems are trained using limited numbers/types of subjects [5,9]. Subjects that are not used to train the neural network model tend to be considered new data domains in machine learning, which can cause performance losses when the model is tested against such domains. Transfer learning is a commonly adopted solution for such issues [12,13]; however, trained models require updating or fine-tuning through lightweight training for transfer learning. New vision data from a new domain (i.e., an untrained subject) is needed for the lightweight training, raising the privacy concern issue once again. Domain discriminators were proposed in recent literature [14,15] to address such domain-adaptive problems. However, they are not effective for classification problems, and their model structure may not be suitable for the data sequence estimation found in human pose tracking tasks.

* Corresponding author.
E-mail addresses: czy0017@auburn.edu (C. Yang), xuyu.wang@csus.edu (X. Wang), smao@ieee.org (S. Mao).

https://doi.org/10.1016/j.dcan.2021.09.002
Received 28 January 2021; Received in revised form 9 August 2021; Accepted 1 September 2021
Available online 13 September 2021

Table 1
Notation.

Symbol	Description
f_s	Frequency of RFID channel s
φ_s	Initial phase offset of RFID channel s
ϕ	RFID phase
η	RFID phase variation
ℓ	Tag-to-antenna distance
H	Input RFID data tensor
n_G	Tag index
n_A	Antenna index
$\eta_{\lfloor t n_G}^{n_A}$	Calibrated phase variation for tag n_G sampled by antenna n_A in time slot t
r, x, y, z	Real numbers in the unit quaternion
i, j, k	Quaternion units in the unit quaternion
$\boldsymbol{\Theta}$	Rotation matrix used in forwarding kinematic
SK	Source initial skeleton
TK	Target initial skeleton
$F_{1:K}$	Input RFID data sequence
$\hat{F}_{1:K}$	Fake RFID data sequence
$\hat{V}_{1:K}$	Estimated skeleton sequence
$V_{1:K}$	Vision data sequence as ground truth
$\mathcal{L}_p$, $\mathcal{L}_c$	Loss functions
$\mathcal{L}_s$, $\mathcal{L}_{all}$	Loss functions
E_{all}	Overall estimation error
$\hat{P}_n$	Estimated 3D position for joint n
$\dot{P}_n$	Ground truth position for joint n

In this paper, we shall address the above challenges encountered while using RFID technology for human pose estimation and propose a novel vision-aided deep-learning-based solution named "Cycle-Pose." The Cycle-Pose system is designed to track the movements of multiple human limbs in real time [1]. With the proposed system, multiple RFID tags are attached to human joints, and their movement is captured by phase variations in RFID response messages upon tag interrogation by a reader. Similar to our prior work "RFID-Pose" [9], Cycle-Pose is a vision-added solution utilizing video data to support a deep learning model in transforming phase variation data to human limb rotations, which differs from traditional tag localization approaches [16]. In addition, we develop a novel deep learning model, i.e., the cycle consistent adversarial network model, to achieve generalization for different subjects. The proposed cycle kinematic network model is trained without the stringent requirements of paired RFID and vision data, allowing the model to transform RFID data into a human skeleton for any subject. We expect our proposed system to achieve better subject generalization compared with traditional schemes when generating human poses for untrained subjects. Cycle-Pose reconstructs 3D human poses in real time from a given initial human skeleton and the rotations of human limbs, which are estimated using the measured RFID phase data. Cycle-Pose only uses video camera data during its training phase and has the desirable advantage of not requiring further vision data in its inference stage, thus preserving user privacy.

We summarize the main contributions of this paper as follows:

- To the best of our knowledge, Cycle-Pose is the first subject-adaptive 3D human pose estimation system to use commodity RFID readers and tags designed to effectively track 3D human poses without using computer vision data in the testing phase.
- We propose a cycle kinematic network model in which the deep learning model is trained using self-supervision. The proposed model learns the transformation using measured RFID phase data to 3D human skeleton for different subjects to achieve effective subject adaptability.
- We propose developing a prototype system with commodity RFID tags/readers, using Kinect 2.0 to obtain the ground truth data required for model training. The proposed system is evaluated with extensive experimentation as well as comparison with a state-of-the-art RFID-Pose system [9]. The experimental results demonstrate that the proposed Cycle-Pose system effectively tracks 3D human poses for different subjects and exhibits great subject adaptability.

In the remainder of this paper, we review related work in Section 2. The proposed Cycle-Pose system is presented in Section 3. Section 4 discusses the challenges and the solutions we propose to address them. Our prototype implementation and experimental performance study are presented in Section 5. Section 6 summarises this paper. The notation used in this paper is summarized in Table 1.

2. Related work

Various sensing systems have been developed to address the human pose estimation problem, such as video cameras, WiFi, and radar. With the development of RFID techniques, RFID tags have been proposed as a promising means of tracking the problem. In this section, we review existing pose estimation works and related RFID sensing systems.

2.1. Traditional pose estimation schemes

Pose estimation was first developed as a computer vision technique using video cameras [17,18]. Deep learning has shown to be highly effective at extracting human skeletons from captured video data, including 2D RGB cameras [19,20], depth cameras [21], and the Vicon system [22]. Among these techniques, Vicon has been shown to achieve the highest accuracy and is commonly used to capture motion in 3D animations and movies. However, such camera-based techniques usually require sufficient lighting and may raise privacy concerns [3].

The fast development of RF sensing systems and related techniques have led to RF signals being utilized to address camera-based system limitations [23]. An RF-based pose tracking system can better preserve user privacy because no vision data is recorded in the device. Moreover, RF-based systems are not limited by lighting conditions [24]. However, unlike video frames, human skeletons cannot be directly extracted from RF data, forcing existing RF-based pose tracking techniques to base themselves on vision-aided training by way of supervised training with labeled vision data. FMCW Radar was originally proposed for leveraging vison-aided training when transferring radar signals into 2D and 3D human poses, utilizing a teacher-student deep learning model [6,24]. MmWave radar was also employed to estimate human skeletons [7]. As another non-intrusive sensor, WiFi devices were also leveraged to estimate 2D and 3D human poses [4,5]; such WiFi and radar-based techniques could effectively track the human pose but were limited by the environmental interference because of their wide propagation range [4,5]. Furthermore, the FMCW radar device incurred a relatively higher cost as it was implemented with the Universal Software Radio Peripherals (USRP) platform.

2.2. RFID-based pose estimation

RFID tags, as low-cost and lightweight sensors, are also promising for human pose tracking. As wearable sensors, the data collected using RFID tags are mainly used to detect subject movement. RFID-based sensing systems are more robust against environmental effects than other RF sensing techniques (e.g., WiFi). Numerous RFID sensing techniques have been developed in recent years, such as vital sign monitoring [25,26], vibration sensing [27], material identification [28], temperature sensing [29], user authentication [30], and anomaly detection [31]. Furthermore, RFID techniques were utilized for indoor localization [16,32–35] and drone navigation [36,37].

Advances in RFID sensing techniques have inspired the development of human pose tracking using RFID tags attached to a human body. The RF-Wear [11] and RF-Kinect systems were proposed to track the movement of particular human limbs. However, these systems have limitations when simultaneously tracking multiple human limbs. To address this issue, RFID-Pose [9] was proposed to track the movement of an entire

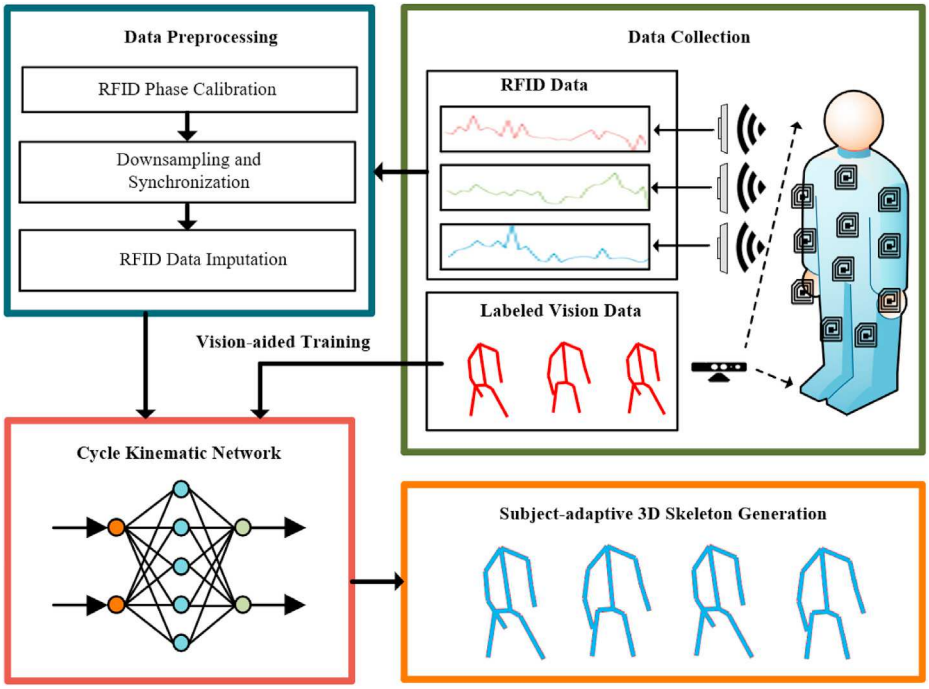

Fig. 1. Cycle-Pose system architecture.

human body and generate the 3D human skeleton in real time. However, the specific paired skeleton training process of RFID-Pose limits user adaptability, and its performance can be affected when it is applied to a new untrained subject. In fact, such a generalization issue is a general challenge for many deep learning-based RF sensing techniques [5,14,15]. Inspired by existing pose tracking systems, we propose a Cycle-Pose system in this paper to improve subject adaptability and overall accuracy in the RFID pose estimation that is currently available with a traditional RFID pose tracking approach.

To the best of our knowledge, the proposed Cycle-Pose system is the first user-adaptive RFID-based 3D human pose tracking system. Cycle-Pose utilizes a novel cross-skeleton training process to address the subject generalization issue, and it is expected to achieve more robust performance than existing techniques when testing different subjects.

3. System overview

In Fig. 1, we present the overall architecture of the proposed Cycle-Pose system, which comprises four modules: (i) Data Collection, (ii) Data Preprocessing, (iii) The Cycle Kinematic Network, and (iv) 3D Skeleton Generation.

3.1. Data collection

The Cycle-Pose system generates a 3D human skeleton using measured RFID phase data. Both RFID data and camera data should be sampled for training the deep learning model. The RFID data is collected from 12 RFID tags attached to human joints, interrogated by three polarized antennas, and used to train the proposed cycle kinematic network. The vision data is collected using Kinect 2.0 on the same subject and action simultaneously as the RFID data is collected. Kinect 2.0 is a depth camera that captures 3D human movements using both an RGB camera and an infrared sensor. 3D movements of each human joint are generated by processing the Kinect data with MATLAB and stored in the form of 3D coordinates for supervised offline training. In the testing phase, the vision data is also used as the ground truth for performance evaluation.

3.2. Data preprocessing

Collected raw RFID data cannot be directly used for 3D skeleton tracking. The RFID collection phase is usually distorted by the RFID communication system's frequency hopping and phase wrapping [8,26,38]. Therefore, it should be calibrated to mitigate the distortion and improve analysis by the proposed neural network model. In addition, because the sampling rates of the RFID reader and Kinect 2.0 are significantly different, the sampled RFID data should be downsampled and synchronized with the vision data. RFID systems use slotted ALOHA-like transmissions; thus, tags are seldom evenly sampled. At most, a single sample exists in a given time slot, and all other RFID phase samples are missing, resulting in sparse RFID data [8]. To address this issue, we employ High Accuracy Low-Rank Tensor Completion (HaLRTC) to estimate the missing RFID data.

3.3. User-adaptive 3D skeleton generation with the cycle kinematic network

We propose a cycle kinematic network to generate 3D poses using calibrated RFID phase data. Unlike traditional RFID-based pose tracking systems [10,11], in which a particular limb's movements are detected, the proposed system detects the 3D coordinates of all human joints simultaneously. Moreover, the proposed cycle kinematic network achieves high subject adaptability, which is not well addressed in prior systems [5,9]. The cycle kinematic network is trained using unpaired RFID data and vision data sampled from different moving subjects. Thus, the trained network can achieve better generalization when transforming RFID data to 3D coordinates for an untrained subject.

4. Challenges and proposed solutions

4.1. RFID phase data calibration

The data's general poor quality is a challenge in human 3D skeleton generation from RFID data. Raw RFID data usually has severe phase distortion and is highly sparse, requiring careful calibration before being used to train a deep learning model. In Fig. 2, we show the RFID

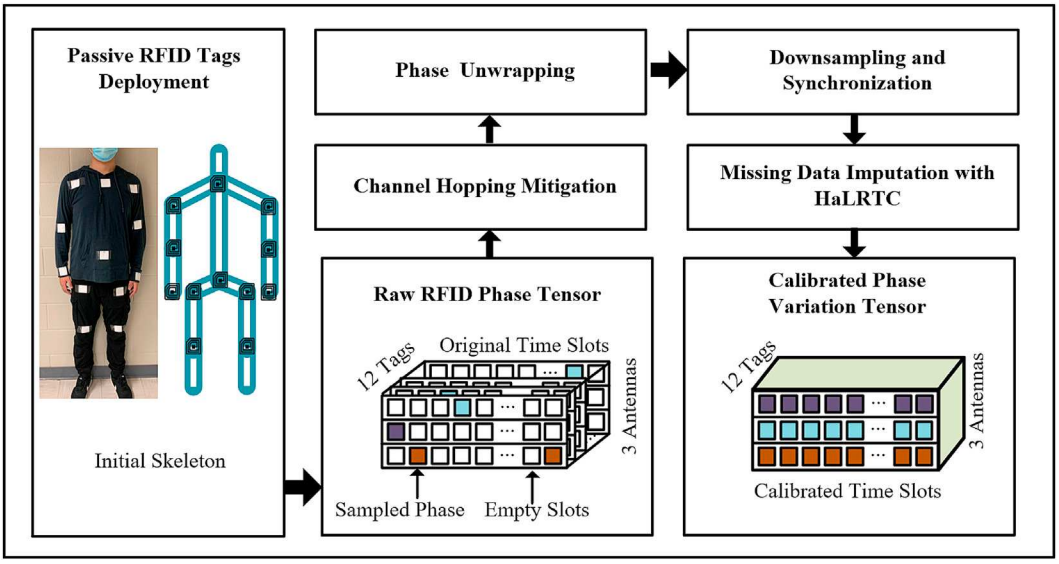

Fig. 2. Flowchart of RFID data preprocessing procedure.

preprocessing procedure in Cycle-Pose, illustrating that passive tags are attached to 12 of the body's joints. RFID phase data is collected by the reader using the low-level protocol when interrogating the tags [26,39].

4.1.1. Phase distortion mitigation

The phase value in a received RFID response indicates the distance between the reader antenna and the tag [39]. Thus, the sampled time series of the RFID phase captures the movement of the RFID tags attached to the subject. The FCC requires channel hopping for RFID communications (i.e., reader-tag channel frequencies must hop among 50 different channels instead of being fixed) [8]. Each time it hops to a new channel, there will be a different phase offset [40]. Consequently, the phase value is dependent on the antenna-tag distance and the current channel used. The phase φ_s sampled on channel s can be written as follows [38]:

$$\varphi_s = \mathrm{mod}\left(\frac{2\pi 2\ell f_s}{c} + \varphi_s^0,\ 2\pi\right),\ s = 1, 2, \ldots, 50 \tag{1}$$

where ℓ is the antenna-tag distance, c is the speed of light, and f_s and φ_s^0 denote the frequency and initial phase offset of channel s, respectively. According to (1), to track the variation of the antenna-tag distance ℓ (i.e., track the tag's movement), the impact of the channel phase offset φ_s^0 should be mitigated first. Fortunately, φ_s^0 is a constant on each channel s. If we use the difference (or *variation*) between two adjacent phase samples on the same channel, the identical channel phase offset components in the two samples will cancel each other. The phase variation η_s^n for channel s is given by

$$\eta_s^n = \mathrm{mod}\left(\frac{4\pi\left(\ell_s^n - \ell_s^{n-1}\right)f_s}{c},\ 2\pi\right),\ s = 1, 2, \ldots, 50,\ n = 2, 3, \ldots \tag{2}$$

where ℓ_s^n represents the antenna-tag distance when the nth sample is measured on channel s. The phase offset φ_s^0 is removed in (2), whereas the displacement of the tag $\ell_s^n - \ell_s^{n-1}$ is retained. The phase distortion due to channel hopping is thus effectively mitigated.

The modulo operations in (1) and (2) also cause considerable phase distortion. Because the collected phase data φ is rounded to the range [0, 2π] rad, sharp phase changes are generated by the modulo operation when the phase crosses 0 rad or 2π rad. To mitigate this rounding error, the phase variation should be unwrapped. Given a 110Hz sampling rate used by the reader, it is reasonable to assume that the phase variation between two consecutive samples should not be greater than π and not less than $-\pi$. We use the following scheme to unwrap the sampled phase variation data η:

$$\eta' = \begin{cases} \eta - 2\pi\eta/|\eta|, & \text{if } |\eta| > \pi \\ \eta, & \text{otherwise} \end{cases} \tag{3}$$

which decides whether the phase variation value should be unwrapped by adding or subtracting 2π. After unwrapping, all sharp phase changes will be smoothed out, and the calibrated phase variation data can effectively represent RFID tag movement.

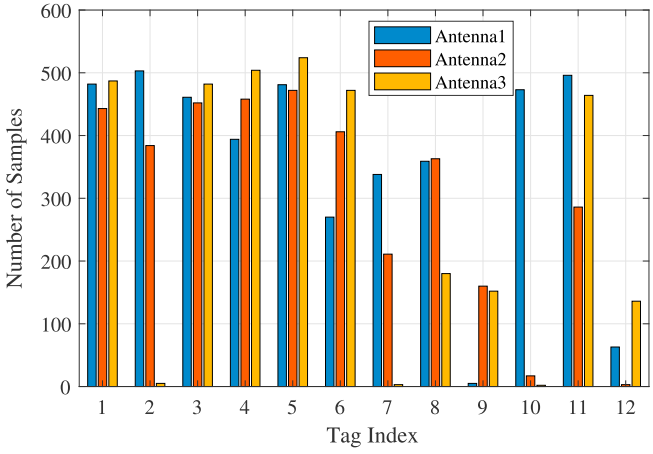

Fig. 3. Illustration of sampling rates for different antennas and tags.

4.1.2. Data imputation

In addition to distortion, the high sparseness of RFID data presents another challenge caused by the Slotted ALOHA-like communications in RFID systems. For each time slot, up to one tag can send its Electronic Product Code (EPC) to a reader, and although the Cycle-Pose system has 12 tags attached to a subject's body, it can also only sample one tag at a time. In the RFID data tensor illustrated in Fig. 2, there is only one sample in each slice (given dimensions of 12 tags × 3 antennas). Thus, the sparsity of the phase data tensor is 35/36, which is considerably high for 3D human pose estimation. Moreover, polarized antennas are used in systems that experience different propagation losses for the tags attached to different human joints. The reflected power from tags with high propagation loss could be too low for antenna detection, causing

significantly different sampling rates. Fig. 3 presents the number of samples collected from the 12 tags using three reader antennas during a period of 200 s. As shown in the figure, there is a considerable divergence in the sampling rates. For example, Tag 5 is frequently read by all the three antennas, Tag 9 is rarely read by Antenna 1, and Tag 10 is rarely read by Antennas 2 and 3. Such imbalanced data greatly affects the performance of 3D human pose tracking that requires data from all 12 tags. To learn the relationship between RFID data and 3D skeleton data obtained from the Kinect, we need to (i) deal with the high sparsity RFID tensor data and (ii) synchronize the RFID phase data (i.e., the input for the deep learning model) with the vision data (i.e., the labeled vision data for supervised training).

To address these problems, we begin by downsampling the RFID data from 110 Hz to 30 Hz to match the 30 fps sampling rate of Kinect 2.0. We then synchronize the RFID and vision data based on simultaneous timestamped samples for both data types collected from the same subject. Notably, we cannot synchronize data from different subjects because the RFID and vision data are not sampled simultaneously in such instances. After synchronization, the RFID phase variation tensor for training the deep learning model can be expressed as:

$$\boldsymbol{H}(:,:,t)=\begin{bmatrix}\eta_{t,1}^{1} & \eta_{t,2}^{1} & \cdots & \eta_{t,n_G}^{1}\\ \eta_{t,1}^{2} & \eta_{t,2}^{2} & \cdots & \eta_{t,n_G}^{2}\\ \vdots & \vdots & \vdots & \vdots\\ \eta_{t,1}^{n_A} & \eta_{t,2}^{n_A} & \cdots & \eta_{t,n_G}^{n_A}\end{bmatrix},\ t=1,2,\ldots,N_t \tag{4}$$

where t indicates the time slot, n_A and n_G are the numbers of antennas and tags, respectively, and $\eta_{t,n_g}^{n_a}$ represents the calibrated phase variation from tag n_g sampled by antenna n_a in time slot t.

As discussed earlier, the high sparsity of the tensors should be addressed to make them useful. However, conventional data interpolation methods (such as bilinear interpolation) are unsuitable for RFID data imputation because of the highly different sampling rates for each tag. Missing samples (such as those of tags 2, 7, and 10 missed by Antenna 3; Fig. 3) require interpolation to make the tensor useful. Because the sampled data from these tags are practically zero, the bilinear interpolation performance is notably poor. However, the samples from different antennas are generated by the same source: tag movement. Thus, although Antenna 3 loses these tags, the lost samples can still be estimated from the samples collected by Antennas 1 and 2. We leverage the tensor completion technique to estimate the missing samples in $\boldsymbol{H}$ (defined in (4)). The algorithm used in the Cycle-Pose system is HaLRTC [38], which can achieve high accuracy in data imputation at a relatively fast speed.

Specifically, data interpolation is accomplished by solving the following optimization problem [41]:

$$\min_{\hat{\boldsymbol{H}}} \ \|\hat{\boldsymbol{H}}\|_* \tag{5}$$

$$\text{s.t.: } \boldsymbol{R}^*\hat{\boldsymbol{H}}=\boldsymbol{R}^*\boldsymbol{H} \tag{6}$$

where $\hat{\boldsymbol{H}}$ is an estimation of the ideal tensor $\boldsymbol{H}_{ideal}$ (comprises all the ideal phase variation data), $\boldsymbol{R}$ is a tensor with the same size as $\boldsymbol{H}$ but with binary elements, $\boldsymbol{R}_{IJK}=1$ when the element $\boldsymbol{H}_{IJK}$ is sampled data, and $\boldsymbol{R}_{IJK}=0$ when the sample is missing. In (5), $\|\cdot\|_*$ denotes the trace norm of tensors, which is related to the singular values of a tensor. Because the number of time slots N_t for each data tensor $\boldsymbol{H}$ is effectively reduced by downsampling, the data imputation process takes less than 0.2 s in the proposed system. We find the HaLRTC algorithm is highly suitable for RFID data imputation because missing data is estimated from the low rank components of the phase variation tensor, which mainly captures tag (human) movements.

4.2. 3D human skeleton generate from RFID data

Most existing human pose tracking systems are based on confidence maps generated from collected signals, such as video cameras [19], WiFi [4], and FMCW radar [6]. Human features are initially captured to form the confidence map, and then a human skeleton can be extracted using the map. However, this technique is unsuitable for RFID-based systems due to the RFID communication protocol's low sampling rate and RFID data samples carrying limited information. For instance, the RFID sampling rate (110 Hz) is significantly lower than that of WiFi (1000 Hz) [5]. Each RFID sample comprises one piece of phase data, whereas WiFi samples each contain phase data from 30 subcarriers (and each sample captured using a video camera comprises a full image). To generate a sequence of confidence maps using captured RFID data at a 10 fps rate, up to 110 RFID phase samples may be obtained per second, and each of that second's 10 frames shall be constructed from 11 RFID phase samples. Even if the map resolution is reduced to 100 × 100, we still need to transform the 11 RFID phase samples to a map with 10, 000 pixels, indicating an obvious *ill-posed problem*. To address this ill-posed problem, we utilize the forward kinematic technique, which has been widely used in robotics and 3D animation [42]. With a given initial skeleton (i.e., the original locations of all the joints and the lengths of all limbs), this technique can estimate the position of each joint based on its relative rotation and the position of its parent joint. For example, when the right elbow position is given, the right-hand position can be calculated using the forearm length and the relative rotation between the hand and elbow.

In the proposed cycle kinematic network, a 3D rotation is represented in the format of a *unit quaternion* based on Ruler's rotation theorem, which is expressed as

$$r+xi+yj+zk \tag{7}$$

where r, x, y, and z are real numbers, and i, j, and k are the quaternion units related to the corresponding three coordinates. Given the 3D position of a joint, represented as $ai+bj+ck$, and a 3D rotation with unit quaternion $r+xi+yj+zk$, the rotation matrix $\boldsymbol{\Theta}$ can be derived as:

$$\boldsymbol{\Theta}=\begin{bmatrix}1-2(y^2+z^2) & 2(xy+zr) & 2(xz-yr)\\ 2(xy-zr) & 1-2(x^2+z^2) & 2(yz+xr)\\ 2(xz+yr) & 2(yz-xr) & 1-2(x^2+y^2)\end{bmatrix} \tag{8}$$

The new updated position of the joint, $a'i+b'j+c'k$, can be calculated as

$$\begin{bmatrix}a'\\ b'\\ c'\end{bmatrix}=\boldsymbol{\Theta}\begin{bmatrix}a\\ b\\ c\end{bmatrix} \tag{9}$$

with the forward kinematic technique, the current human pose is determined based on the previous human pose and the 3D rotation of each joint. To estimate the new positions of the 12 human joints, only 48 parameters need to be determined from the RFID samples. Compared with the traditional approach of generating a 10, 000-pixel map, the proposed technique effectively reduces problem complexity and can achieve improved accuracy as well.

4.3. Achieve subject generalization

Although the forward kinematic technique can effectively address the ill-posed problem, the subject's initial skeleton is still required, limiting the model's adaptability to untrained subjects. People have different skeleton forms, so to ensure that the deep learning model can successfully generate 3D skeletons for different subjects, the training dataset should include all types of human skeletons, leading to a high cost for collecting labeled data. If the network is trained with a limited amount of skeletons, model performance might suffer when testing a subject possessing a skeleton not included in the training dataset [5]. This is because the

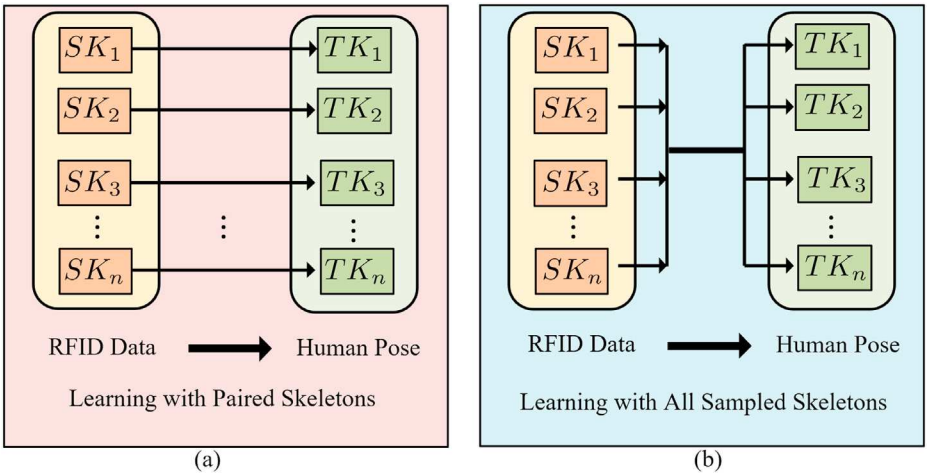

Fig. 4. Different methodologies for deep learning model training. (a) training with paired skeletons; (b) training with all sampled skeletons.

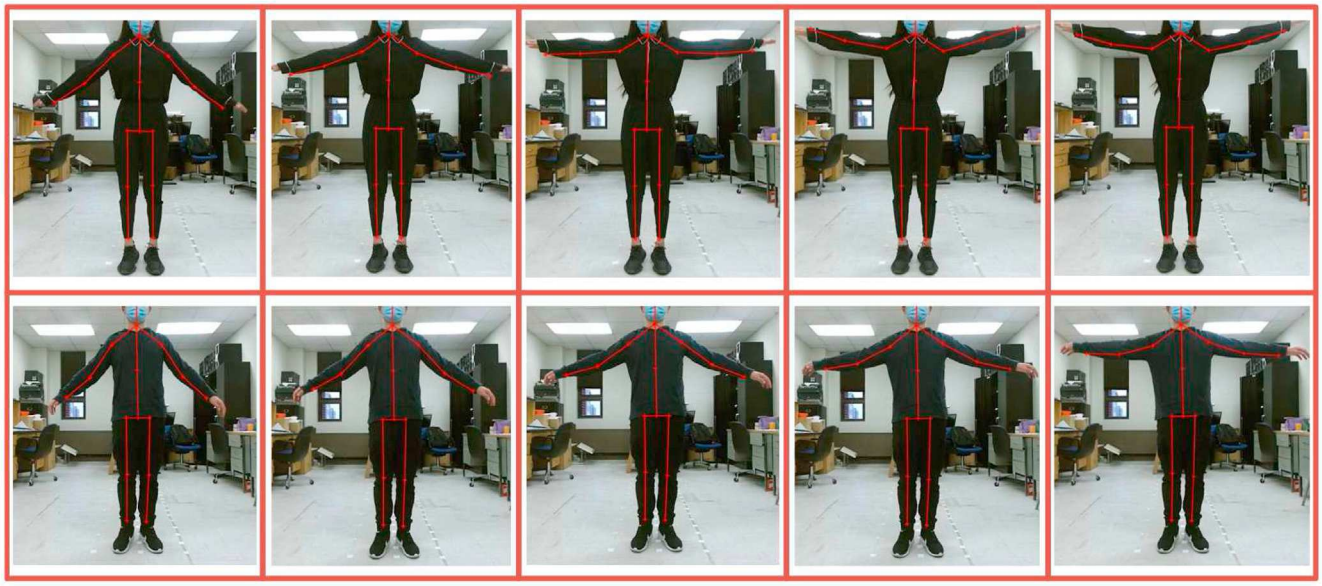

Fig. 5. Labeled pose data sampled by Kinect 2.0 for two different subjects. The upper row is for Subject 1 and the lower row is for Subject 2.

traditional training process is performed with the same source initial skeleton and target initial skeleton, as illustrated by Fig. 4(a). In the plot, SK_n represents the source initial skeleton for subject n from the RFID data, and TK_n represents the target initial skeleton from the vision data with $TK_n = SK_n$. The traditional training methodology aims to learn the relationship between 3D skeleton coordinates and the RFID data for the same skeleton. Therefore, the training results will be suitable for these n specific skeletons included in the training data, but the well-trained model may not perform well when used to test a new skeleton.

4.3.1. *Cross-skeleton learning methodology*

To achieve high subject generalization, the deep learning model must learn the relationship between different source and target skeletons, effectively transforming RFID data to 3D skeletons regardless of the subject skeleton being part of the training dataset. In this paper, we propose a new training methodology, as shown in Fig. 4(b). With this methodology, the training focuses on learning with paired skeletons as well as different source and target skeletons. For example, a specific movement type (e.g., kicking) would utilize all RFID and vision data during training, even if the movement data was not sampled from the same subject. Thus, the network will learn how to transfer RFID data to 3D human poses with different initial skeletons (e.g., SK_1, SK_2, …, and SK_n). Because the network is not trained with a specific initial skeleton, the well-trained model can achieve higher subject adaptability compared with a traditional network structure.

Unfortunately, training with different *SK*s and *TK*s is challenging owing to the significant variance in the training data of two different subjects. Despite performing identical movements, different subjects can exhibit differing speeds and scales. This is illustrated in Fig. 5, which shows two subjects with differing limb lengths and possibly different RFID tag positions. Fig. 5 shows the skeletons obtained using Kinect 2.0 when two different subjects perform the same action (i.e., waving arms), sampled at the same frame rate. The speeds and arm movement ranges differ between subjects.

The considerable disparity illustrated in Fig. 5 indicates that the deep learning model will not be easily trained, considering the position loss between estimated and ground-truth poses for different subjects. A self-supervised network is therefore required for cross-skeleton learning with unpaired initial skeletons.

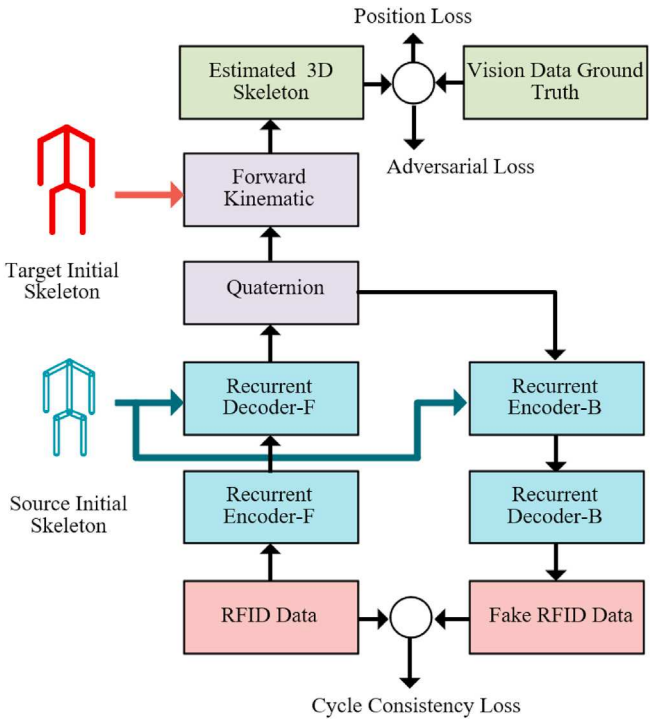

Fig. 6. Overview of the proposed cycle kinematic network model.

4.3.2. Proposed cycle kinematic network

A cycle-consistent adversarial network is an advanced neural network model initially designed for image-to-image translation with unpaired training datasets [43]. Cycle consistent networks have also been used to temporally align the frames of two different video streams captured for the same event [44]. Such networks can generate fake input data from the output data, allowing the network to train using self-supervision with real and fake generated input data. Thus, this training method decreases the need for paired labeled data expected by traditional neural networks. Motivated by the advantages of a cycle consistent adversarial network, we develop a cycle kinematic network to address the technical challenges of cross-skeleton learning. Our proposed model structure (presented in Fig. 6) extracts RFID phase variation sequence features using a recurrent encoder, named "recurrent encoder-forward" or "recurrent encoder-F." With additional input on a subject's source initial skeleton, the recurrent decoder-Forward (recurrent decoder-F) translates human movement features captured by RFID phase variations into the forms of unit quaternion, which represent the 3D rotation of the subject's joints. The unit quaternion is then utilized using the forward kinematic algorithm to generate a 3D human skeleton with a given target initial skeleton. The cycle consistent network is leveraged to recover the RFID data from the estimated quaternion, which is constructed using the recurrent encoder-backward (recurrent encoder-B) and recurrent decoder-backward (recurrent decoder-B). If the translation from RFID phase variation to 3D limb rotation is successful, the inverse transformation will be obtained using recurrent encoder-B and decoder-B.

With the fake generated RFID data, the model can be trained with self-supervision, and the ground truth provided by the vision data does not need to be strictly paired with the RFID input data. Although both recurrent decoder-F and encoder-B are trained through the self-supervised training process, only recurrent decoder-F is leveraged for skeleton generation in the testing phase. Thus, the proposed cycle kinematic network does not incur additional overhead for human skeleton tracking than its predecessor [9].

4.3.3. Loss function design

The loss function used to train the proposed cycle kinematic network comprises three parts, including position, adversarial, and cycle consistency loss (Fig. 6). We also leverage smoothing loss to further improve the tracking accuracy. When there are K training steps, we define the calibrated RFID phase variation sequence as $F_{1:K}$ and the reconstructed fake RFID data as $\hat{F}_{1:K}$. The estimated skeleton from the neural network is denoted by $\hat{V}_{1:K}$, and the vision data sequence used for supervision is denoted by $V_{1:K}$. The position loss is defined as the difference between the estimated 3D skeleton and the ground truth given by

$$\mathcal{L}_p = \|\hat{V}_{1:K} - V_{1:K}\|_2^2 \tag{10}$$

In other words, position loss is determined by the 3D position divergence between the estimated pose data and the ground truth from the Kinect data.

For unpaired training data collected from different skeletons, $\mathcal{L}_p$ also includes the error caused by asynchronous datasets. We define the cycle consistency loss as

$$\mathcal{L}_c = \|\hat{F}_{1:K} - F_{1:K}\|_2^2 \tag{11}$$

which represents the difference between the input RFID data $F_{1:K}$ and the fake RFID data $\hat{F}_{1:K}$ generated by the cycle consistent network. The effect of unpaired data can be mitigated by $\mathcal{L}_c$ in the overall loss function.

We also consider the smooth loss for improving the accuracy of the cycle pose network. As shown in (10), the position loss is independent of each timestamp. However, because human movement is a smooth continuous process, estimated positions should only change slightly between two adjacent sequence data (i.e., no big sudden jumps). To ensure that consecutive changes in each time step are bounded, we propose the smooth loss $\mathcal{L}_s$, which is given by

$$\mathcal{L}_s = \|\hat{V}_{2:K+1} - \hat{V}_{1:K}\|_2^2 + \|\hat{F}_{2:K+1} - \hat{F}_{1:K}\|_2^2 \tag{12}$$

The overall loss function for generator network G is defined as a weighted sum of the three losses:

$$\mathcal{L}_{all} = Q_p\mathcal{L}_p + Q_c\mathcal{L}_c + Q_s\mathcal{L}_s \tag{13}$$

G can be effectively trained using the overall loss function $\mathcal{L}_{all}$ for consecutive human pose tracking regardless of the subject used to sample RFID and vision data. In this paper, we set $Q_p = 0.6$, $Q_c = 0.4$, and $Q_s = 0.02$ according to our experimental results.

The adversarial loss is defined to determine if the network is well trained or not, which is represented by a realism score calculated using a discriminator network D [42]:

$$\mathcal{L}_D = D(\hat{V}_{2:K} - \hat{V}_{1:K-1}, V_{2:K} - V_{1:K-1}) \tag{14}$$

Eq. (14) shows that the discriminator's input is not the position loss but the variation between the previous and current skeletons in V and $\hat{V}$, respectively. Although V and $\hat{V}$ are unpaired data sequences, the discriminator can determine whether the movements performed by the two subjects are identical. This is because, for the same type of movement, all joint variations between two adjacent data sequences should be similar, regardless of the two subject's movements being synchronized. The overall training objective of the generative adversarial network can be represented as

$$\hat{G} = arg\,min_G \max_D \mathcal{L}_D \tag{15}$$

where $\hat{G}$ represents the target generator network (i.e., recurrent encoder-F and decoder-F). We set a realism score threshold to balance the discriminator D and the generator G. When G can successfully fool D, and

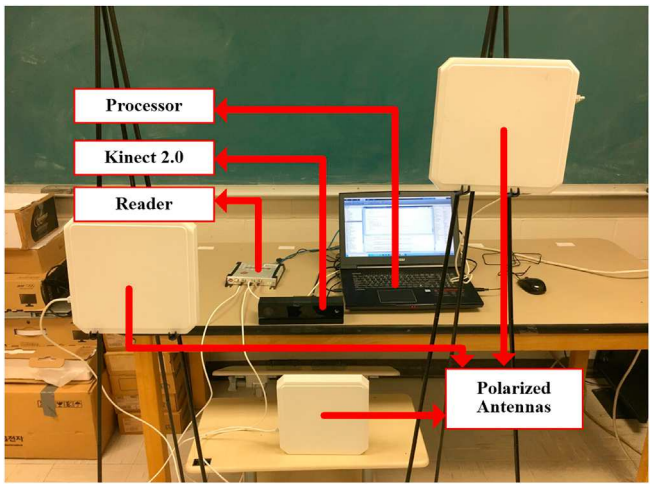

Fig. 7. Hardware configuration of the Cycle-Pose prototype system.

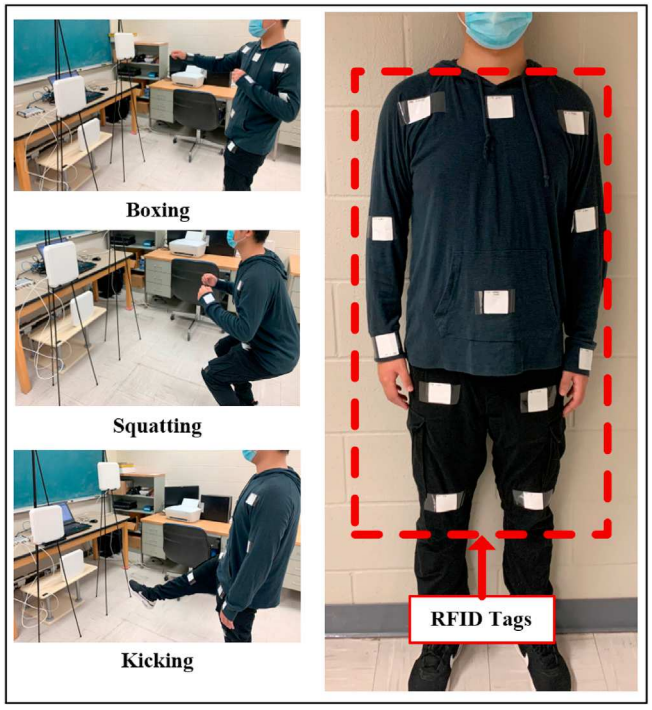

Fig. 8. RFID tag deployment and motion sampling.

the network will be well trained and be able to effectively transform RFID data into 3D skeletons.

5. System prototype and experimental study

5.1. Cycle-Pose implementation

To evaluate the performance of Cycle-Pose, we developed a prototype system using an off-the-shelf Impinj R420 reader equipped with three S9028PCR polarized antennas (Fig. 7). The RFID tags used in Cycle-Pose are ALN-9634 (HIGG-3). The vision data, used for training supervision as well as the ground truth for test accuracy evaluation, was collected using a Kinect 2.0 device similar to prior works [10]. We have found that Kinect's accuracy is affected by camera angle, lighting conditions, and the distance to the subject. To minimize Kinect data errors, we constrained subject positions and provided sufficient lighting during data collection.

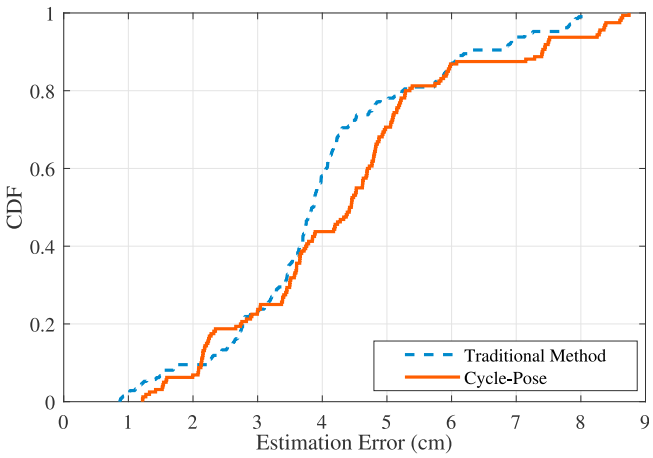

Fig. 9. CDF curves of the two schemes when testing three trained subjects.

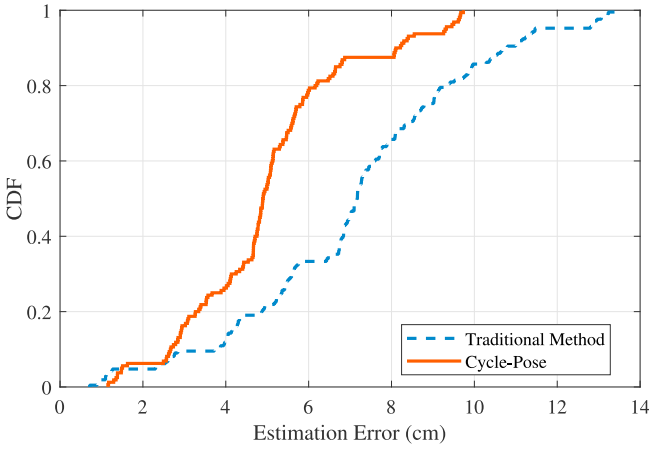

Fig. 10. CDF curves of the two schemes when testing two untrained subjects.

The Kinect system has a much lower cost than commercial human motion capture systems while providing the required ground truth with an acceptable error [10].

As shown in Figs. 8 and 12 RFID tags were attached to the following subject joints: left shoulder, left elbow, left wrist, right shoulder, right elbow, right wrist, neck, pelvis, left hip, left knee, right hip, and right knee. Head and feet were omitted in our prototype system because of the limited scanning range of the RFID antennas used by Cycle-Pose. Although more antennas could be used to scan a subject's entire body, constructing a skeleton with 12 joints is sufficient for monitoring most human activities. The three antennas placed at different heights ensured that all RFID tags could be scanned by at least one antenna.

Cycle-Pose collected RFID phase data when subjects were repeatedly performing specific motions in front of the three antennas. Two types of motion were sampled to train the cycle kinematic network: (i) Simple motions that involved the movement of a single limb; (ii) Compound motions comprising entire body movement (e.g., boxing, walking, drinking, body twisting, and deep squatting).

We conducted extensive experimentation on five volunteer subjects (four male, and one female) to evaluate the Cycle-Pose system's performance, particularly its subject generalization capability. Each subject had a different age, weight, and height, introducing five different initial skeletons. All subjects were required to wear long shirts and pants for data collection, with tags attached to their clothes. The network was trained with only three subjects, while the remaining two subjects were involved in testing. Considering the variety of initial skeletons, subject

Table 2
Estimation errors for different subjects.

Subject	Estimation Error (Baseline)/cm	Estimation Error (Cycle-Pose)/cm
Subject 1 (trained)	3.72	4.12
Subject 2 (trained)	4.55	4.43
Subject 3 (trained)	3.58	3.79
Subject 4 (untrained)	5.32	4.51
Subject 5 (untrained)	8.17	4.97

generalization could be demonstrated by the tracking accuracy of untrained subjects.

5.2. Experimental results and analysis

For comparative research, the traditional method (RFID-Pose) was used as a baseline scheme [9], which trains its network with paired RFID and vision data. Identical datasets were used to train and test both methods; the overall accuracy results are presented in Fig. 9 and Fig. 10. The overall estimation error $\mathcal{E}_{all}$ used in our experimental evaluation was calculated between the estimated 3D pose data and the ground truth vision data:

$$\mathcal{E}_{all} = \frac{1}{12}\sum_{n=1}^{12} \|\hat{P}_n - \dot{P}_n\| \tag{16}$$

where $\hat{P}_n$ denotes the estimated position of joint n, $\dot{P}_n$ is the ground truth position collected by Kinect 2.0 for joint n in the 3D space, and $\|\hat{P}_n - \dot{P}_n\|$ is the Euclidean distance between the two 3D positions.

5.2.1. Comparison with the baseline scheme

The mean estimation errors for all subjects are presented in Table 2. For the traditional pose estimation technique, the estimation errors for the two untrained subjects were considerably higher than that of the three trained subjects, especially for Subject 5. By contrast, the Cycle-Pose system produced similar estimation errors for all five subjects regardless of training.

Fig. 9 presents the Cumulative Distribution Functions (CDFs) of the estimation errors for both RFID-Pose and Cycle-Pose when the first three subjects were involved in the training process and tested. The CDF curves indicate a median estimation error for the traditional method (i.e., RFID-Pose) of 3.83 cm, whereas the median error for Cycle-Pose was 4.44 cm. The maximum error for Cycle-Pose was 8.64 cm, slightly higher than the traditional method (8.09 cm). These results show that the accuracy of Cycle-Pose was slightly lower than that of the traditional method when testing a trained skeleton because the Cycle-Pose system not only learns the translation from RFID data to a 3D skeleton, but also learns the transformation from different source skeletons to target skeletons. Although the additional learning task affects system performance for specific skeletons, the accuracy degradation is small and acceptable for most skeleton tracking applications (e.g., video gaming and human motion recognition).

The advantages of Cycle-Pose become clear when testing with untrained subjects. Figs. 11 and 12 illustrate a comparison between the two schemes when an untrained subject is squatting and walking, respectively. These two figures indicate that skeletons reconstructed by Cycle-Pose were highly similar to their corresponding ground truths, whereas traditionally generated skeletons exhibited higher estimation errors.

The CDF curves for the two untrained subjects (using the models trained by the first three subjects) are plotted in Fig. 10. Both systems showed increased median and maximum estimation errors when compared with the trained subjects in Fig. 9 (Cycle-Pose: 4.88 cm median, 9.73 cm maximum; RFID-Pose: 7.66 cm median, 12.23 cm maximum). The traditional model was only trained by paired RFID and vision data for the same subject. The training domain was restricted to the specific initial skeleton. When testing with an untrained subject with a different initial skeleton, the traditional model exhibited poorer subject adaptability than Cycle-Pose. In summary, although the accuracy of Cycle-Pose was slightly lower than that of the traditional RFID pose tracking technique when testing with a known subject, the proposed model achieved high subject adaptability when testing untrained subjects.

5.2.2. Accuracy for different joints and motions

Fig. 13 presents the estimation error for each tagged joint; and the

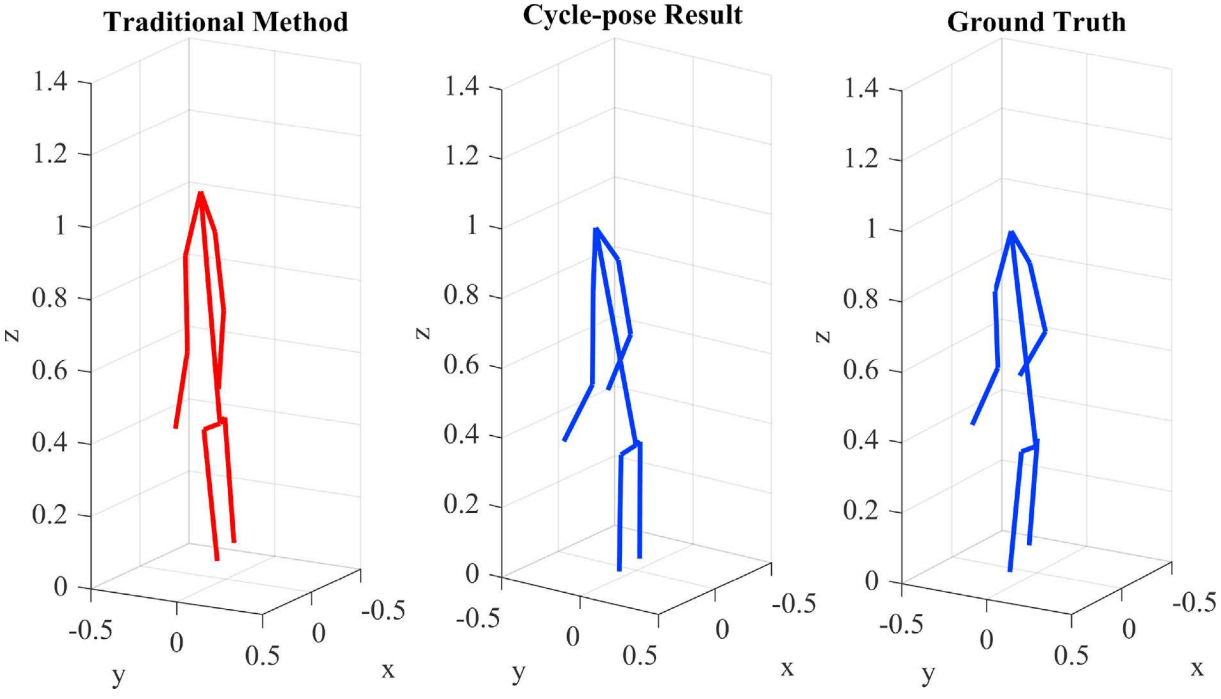

Fig. 11. Comparison results when untrained subject is squatting.

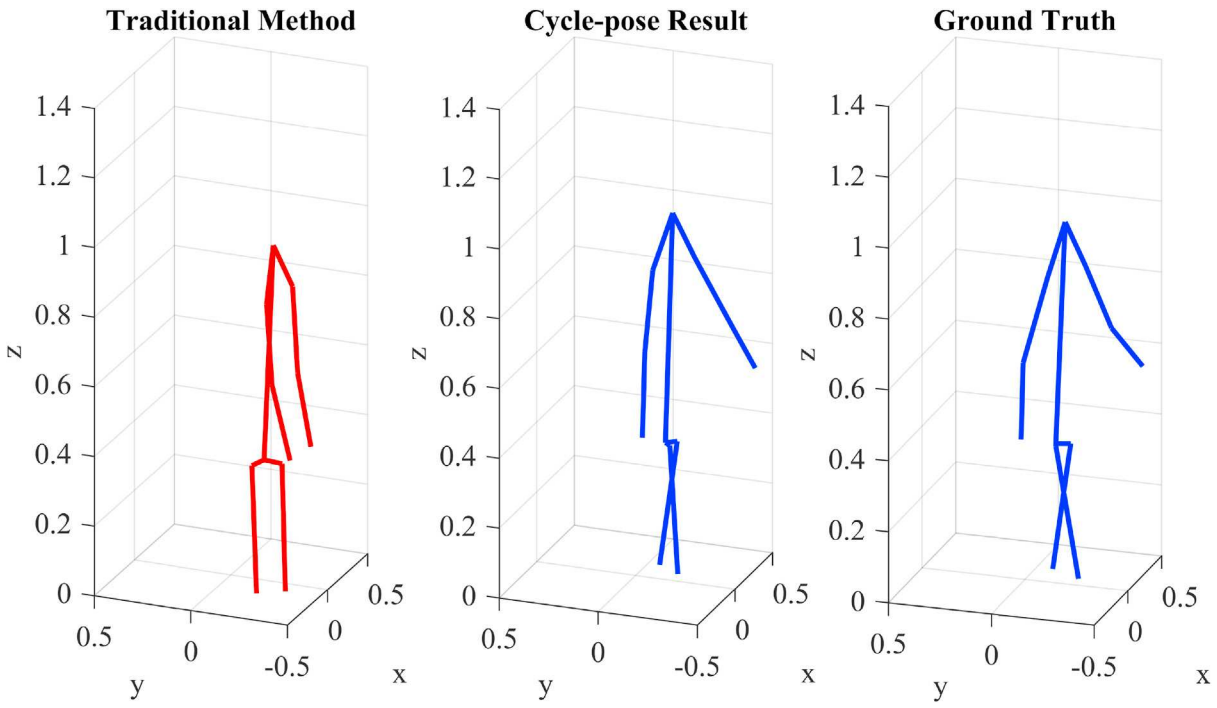

Fig. 12. Comparison results when untrained subject is walking.

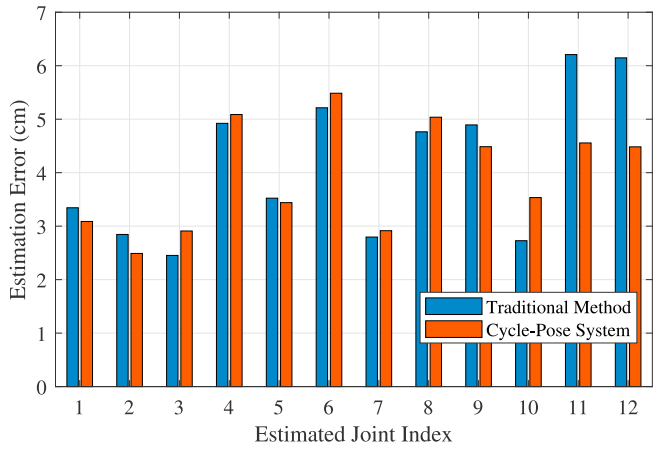

Fig. 13. Estimation error for each human joint, including pelvis, neck, left hip, left knee, right hip, right knee, left shoulder, left elbow, left wrist, right shoulder, right elbow, and right wrist (indexed from 1 to 12).

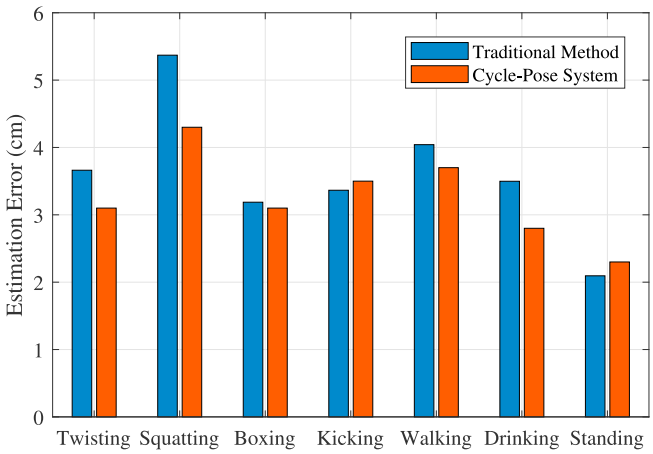

Fig. 14. Estimation errors for tracking different motions.

joints were numbered sequentially from 1 to 12 as follows: pelvis, neck, left hip, left knee, right hip, right knee, left shoulder, left elbow, left wrist, right shoulder, right elbow, and right wrist. We found that estimation errors for both systems were higher than 4.1 cm for the knees, elbows, and wrists. This relatively higher set of errors was mainly because of the forward kinematic technique. When calculating a joint's position based on its parent joint, previous joint errors will accumulate. Thus, the estimation error of the torso will affect the accuracy of the limbs. Fig. 13 also shows that Cycle-Pose achieved higher accuracy than the traditional method when tracking the wrists (i.e., joints 9 and 12). This is due to the Cycle-Pose system performing better with cross-skeleton training when testing different subjects. However, the traditional method's joint estimation accuracy was affected by the untrained subjects, particularly when tracking the wrists.

To evaluate performance for different motions, we plotted the estimation error for each specific movement in Fig. 14, including body twisting, squatting, drinking, walking, kicking, boxing, and standing still. As the figure shows, the pose estimation accuracy differed for different motions; the largest errors occur when tracking deep squatting (5.27 cm for RFID-Pose and 4.33 cm for Cycle-Pose). These errors can be attributed mainly to the pelvis joint errors. As a root joint of the human skeleton, pelvis position estimation does not benefit from the forward kinematic technique and the smooth loss function $\mathcal{L}_s$. Therefore, when testing the motion of frequent pelvis movements, overall accuracy will be degraded. Nevertheless, the Cycle-Pose system achieved lower estimation errors compared with the traditional system for most motions.

5.2.3. *Impact of the effective sensing range*

We also evaluated the impact of the effective sensing range of the Cycle-Pose system in terms of the sampling rate of all tags of each antenna. Fig. 15 shows the sampling rates (i.e., measured samples by each antenna at each second) when the three antennas were placed at different distances from the subject. As the figure shows, distances longer than 5.5 m caused the sampling rates for all antennas to decrease below 60 per second, the lower threshold for effective human pose tracking in real

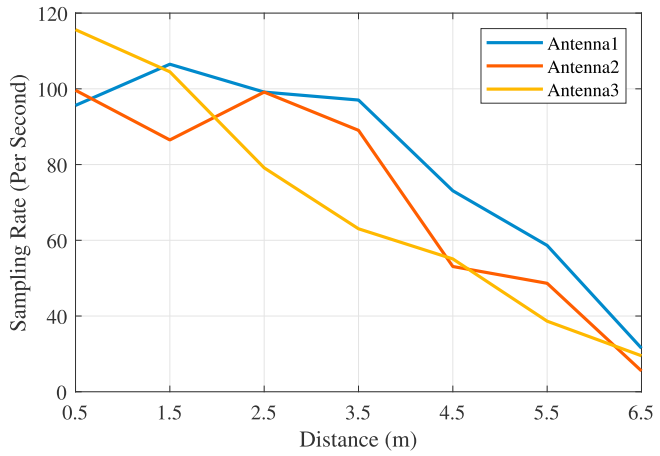

Fig. 15. Sampling rates of the three antennas when the antennas are placed at different distances from the subject.

time. Therefore, we conclude that the Cycle-Pose system should be deployed within a range of 4.5 m for effective sensing.

6. Conclusions

In this paper, we proposed a subject-adaptive, real-time 3D pose estimation and tracking system called Cycle-Pose. A preprocessing module was proposed to effectively mitigate the effect of phase distortion and missing RFID data samples. The proposed system then leveraged a novel cycle kinematic network to estimate human postures in real time using RFID phase data, which was trained with unpaired RFID and vision data sampled from different subjects. The Cycle-Pose system was implemented with commodity RFID tags/reader and compared using a traditional RFID based technique, i.e., RFID-Pose, in our experimental study. Its high subject adaptability and accuracy were demonstrated in our experimental study using Kinect 2.0 as ground truth. Our study also provided useful insights to improve the generalization of other deep learning-based RF sensing systems.

Acknowledgments

This work is supported in part by the US National Science Foundation (NSF) under Grants ECCS-1923163 and CNS-2107190, and through the Wireless Engineering Research and Education Center at Auburn University.

References

[1] C. Yang, X. Wang, S. Mao, Subject-adaptive skeleton tracking with RFID, in: Proc. IEEE MSN 2020, IEEE, 2020, pp. 599–606.
[2] M. Andriluka, S. Roth, B. Schiele, Monocular 3D pose estimation and tracking by detection, in: Proc. IEEE CVPR IEEE, 2010, pp. 623–630.
[3] Tom's Guide, Millions of Wireless Security Cameras Are at Risk of Being Hacked: what to Do, 2020. https://www.tomsguide.com/news/hackable-security-cameras, 2020 (accessed Aug. 28, 2020).
[4] F. Wang, S. Zhou, S. Panev, J. Han, D. Huang, Person-in-WiFi: fine-grained person perception using WiFi, in: Proc. IEEE ICCV 2019, Republic of Korea, Seoul, 2019, pp. 5452–5461.
[5] W. Jiang, H. Xue, C. Miao, S. Wang, S. Lin, C. Tian, S. Murali, H. Hu, Z. Sun, L. Su, Towards 3D human pose construction using WiFi, in: Proc. ACM MobiCom'20, 2020, pp. 1–14.
[6] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, D. Katabi, Through-wall human pose estimation using radio signals, in: Proc. IEEE CVPR IEEE, 2018, pp. 7356–7365.
[7] A. Sengupta, F. Jin, R. Zhang, S. Cao, mm-Pose: real-time human skeletal posture estimation using mmWave radars and CNNs, IEEE Sensor. J. 20 (17) (2020) 10032–10044.
[8] J. Zhang, S. Periaswamy, S. Mao, J. Patton, Standards for passive UHF RFID, ACM GetMobile 23 (3) (2019) 10–15.
[9] C. Yang, X. Wang, S. Mao, Rfid-Pose: Vision-aided 3D Human Pose Estimation with RFID, IEEE Transactions on Reliability. 70 (3) (2021) 1218–1231.
[10] C. Wang, J. Liu, Y. Chen, L. Xie, H.B. Liu, S. Lu, Rf-Kinect, A wearable RFID-based approach towards 3D body movement tracking, Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol. 2 (1) (2018) 1–28.
[11] H. Jin, Z. Yang, S. Kumar, J.I. Hong, Towards wearable everyday body-frame tracking using passive RFIDs, Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol. 1 (4) (2018) 1–23.
[12] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2009) 1345–1359.
[13] R. Raina, A. Battle, H. Lee, B. Packer, A.Y. Ng, Self-taught learning: transfer learning from unlabeled data, in: Proc. ACM ICML 2007, Corvalis, OR, 2007, pp. 759–766.
[14] W. Jiang, et al., Towards environment independent device free human activity recognition, in: Proc. ACM MobiCom ACM, 2018, pp. 289–304.
[15] Wang, Fangxin, Jiangchuan Liu, and Wei Gong, Multi-adversarial in-car activity recognition using RFIDs, IEEE Transactions on Mobile Computing. 20(6) (2020): 2224–2237.
[16] C. Yang, X. Wang, S. Mao, SparseTag, High-precision backscatter indoor localization with sparse RFID tag arrays, in: Proc. IEEE SECON IEEE, 2019, pp. 1–9.
[17] Y. Chen, Y. Tian, M. He, Monocular human pose estimation: a survey of deep learning-based methods, Elsevier Computer Vision and Image Understanding 192 (3) (2020) 1–20.
[18] R. Mitra, N.B. Gundavarapu, A. Sharma, A. Jain, Multiview-consistent semi-supervised learning for 3D human pose estimation, in: Proc. IEEE CVPR IEEE, 2020, pp. 6907–6916.
[19] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, in: Proc. IEEE CVPR IEEE, 2017, pp. 7291–7299.
[20] X. Fan, K. Zheng, Y. Lin, S. Wang, Combining local appearance and holistic view: dual-source deep neural networks for human pose estimation, in: Proc. IEEE CVPR IEEE, 2015, pp. 1347–1355.
[21] Z. Zhang, Microsoft Kinect sensor and its effect, IEEE Multimedia 19 (2) (2012) 4–10.
[22] L. Sigal, A.O. Balan, M.J. Black, Humaneva, Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, Springer Int. J. Computer Vision 87 (1/2) (2010) 1–24.
[23] X. Wang, X. Wang, S. Mao, RF sensing for Internet of Things: a general deep learning framework, IEEE Commun. Mag. 56 (9) (2018) 62–69.
[24] M. Zhao, Y. Tian, H. Zhao, M.A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, A. Torralba, RF-based 3D skeletons, in: Proc. ACM SIGCOM ACM, 2018, pp. 267–281.
[25] Y. Hou, Y. Wang, Y. Zheng, Tagbreathe: monitor breathing with commodity RFID systems, in: Proc. IEEE ICDCS IEEE, 2017, pp. 404–413.
[26] C. Yang, X. Wang, S. Mao, Unsupervised drowsy driving detection with RFID, IEEE Trans. Veh. Technol. 69 (8) (2020) 8151–8163.
[27] P. Li, Z. An, L. Yang, P. Yang, Towards physical-layer vibration sensing with RFIDs, in: Proc. IEEE INFOCOM 2019, France, Paris, 2019, pp. 892–900.
[28] J. Wang, J. Xiong, X. Chen, H. Jiang, R.K. Balan, D. Fang, TagScan: simultaneous target imaging and material identification with commodity RFID devices, in: Proc. ACM MobiCom 2017, Snowbird, Utah, 2017, pp. 288–300.
[29] X. Wang, J. Zhang, Z. Yu, S. Mao, S. Periaswamy, J. Patton, On remote temperature sensing using commercial UHF RFID tags, IEEE Internet of Things Journal 6 (6) (2019) 10715–10727.
[30] C. Zhao, Z. Li, T. Liu, H. Ding, J. Han, W. Xi, R. Gui, Rf-Mehndi, A fingertip profiled RF identifier, in: Proc. IEEE INFOCOM 2019, France, Paris, 2019, pp. 1513–1521.
[31] J. Guo, T. Wang, Y. He, M. Jin, C. Jiang, Y. Liu, Twinleak: RFID-based liquid leakage detection in industrial environments, in: Proc. IEEE INFOCOM 2019, France, Paris, 2019, pp. 883–891.
[32] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, Y. Liu, Tagoram: real-time tracking of mobile RFID tags to high precision using COTS devices, in: Proc. ACM MobiCom 2014, Maui, HI, 2014, pp. 237–248.
[33] J. Wang, D. Katabi, Dude, where's my card? RFID positioning that works with multipath and non-line of sight, in: Proc. ACM SIGCOMM'13, ACM, 2013, pp. 51–62.
[34] L. Shangguan, K. Jamieson, The design and implementation of a mobile RFID tag sorting robot, in: Proc. ACM MobiSys 2016, 2016, pp. 31–42. Singapore.
[35] Y. Ma, N. Selby, F. Adib, Minding the billions: ultra-wideband localization for deployed RFID tags, in: Proc. ACM MobiCom 2017, Snowbird, Utah, 2017, pp. 248–260.
[36] J. Zhang, Z. Yu, X. Wang, Y. Lyu, S. Mao, S.C. Periaswamy, J. Patton, X. Wang, RFHUI: an intuitive and easy-to-operate human-uav interaction system for controlling a UAV in a 3D space, in: Proc. EAI MobiQuitous EAI, 2018, pp. 69–76.
[37] J. Zhang, X. Wang, Z. Yu, Y. Lyu, S. Mao, S. Periaswamy, J. Patton, X. Wang, Robust RFID based 6-DoF localization for unmanned aerial vehicles, IEEE Access J 7 (1) (2019) 77348–77361.
[38] C. Yang, X. Wang, S. Mao, Respiration monitoring with RFID in driving environments, IEEE J. Sel. Area. Commun. 39 (2) (2021) 500–512.
[39] M. Lenehan, Application note - low level user data support [online] Available: https://support.impinj.com/hc/en-us/articles/202755318-Application-Note-Low-Level-User-Data-Support, 2019 (accessed Aug. 28, 2020).
[40] C. Yang, X. Wang, S. Mao, Unsupervised detection of apnea using commodity RFID tags with a recurrent variational autoencoder, IEEE Access Journal 7 (1) (2019) 67526–67538.
[41] J. Liu, P. Musialski, P. Wonka, J. Ye, Tensor completion for estimating missing values in visual data, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2012) 208–220.
[42] R. Villegas, J. Yang, D. Ceylan, H. Lee, Neural kinematic networks for unsupervised motion retargetting, in: Proc. IEEE CVPR IEEE, 2018, pp. 8639–8648.
[43] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: ICCV IEEE, 2017, pp. 2223–2232.
[44] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, A. Zisserman, Temporal cycle-consistency learning, in: Proc. IEEE CVPR IEEE, 2019, pp. 1801–1810.