2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall) | 978-1-6654-5468-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/VTC2022-Fall57202.2022.10013052

Data Augmentation for RFID-based 3D Human Pose Tracking

Ziqi Wang, Chao Yang, and Shiwen Mao
Dept. of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201
Email: {zzw0104, czy0017}@auburn.edu, smao@ieee.org

Abstract—**Interest in Radio Frequency (RF) based 3D human pose tracking has skyrocketed in the age of Artificial Intelligence of Things (AIoT). Compared to Computer Vision (CV) based methods, RF-based approaches are more resilient to lighting and non-line-of-sight conditions, and can better preserve user privacy. However, the majority of the current RF-based methods rely on a vision-aided multi-modal learning approach. An extensive amount of paired training data, i.e., Radio-Frequency Identification (RFID) data and vision data, must be collected, to achieve an adequate performance with the supervised-learning network. In order to mitigate such time-consuming and costly tasks, we propose a data augmentation method based on Generative Adversarial Network (GAN), named RFPose-GAN, to generate synthesized RFID data to alleviate the complications of using commodity RFID tags and receivers. In this paper, a forward kinematic layer is incorporated to generate simulated vision pose data, thus eliminating the need of using a Kinect 2.0 device in RFPose-GAN. Experiments conducted demonstrate that the synthesized data achieves accurate pose estimation performance.**

I. Introduction

Estimating the human pose is an important task with various applications in, e.g., video surveillance, gaming, and activity recognition, etc. Nonetheless, occlusion continues to be a major challenge in computer vision (CV)-based techniques due to the camera angle, obstacles in the line-of-sight path, and the illumination condition. Furthermore, the CV-based datasets are prone to malicious attacks, which raises increasing security and privacy concerns.

To address the privacy issue, our previous work RFID-Pose [1] utilized RFID for its stronger robustness to environment interference than other RF sensing based solutions, such as WiFi and Frequency Modulated Continuous Wave (FMCW) radar. In this system, to capture the body movements in RFID phase data, 12 passive RFID tags are attached to the correspondent joints of the subject. The paired and synchronized data sequences captured concurrently by a Kinect 2.0 device and an RFID reader are used to train a deep kinematic neural network. After the training, RFID-Pose can rebuild a 3D human pose in real time with just RFID data, eliminating the requirement for visual input. Although RFID-Pose achieves an excellent performance, the training process necessitates a significant amount of paired RFID and Kinect data. The data collection process is quite time-consuming, requiring the test subject to perform a variety of tasks for an extended period of time in front of the Kinect camera and RFID reader, which motivated us to cut down the cost of the data collection.

In our prior work RFPose-GAN [2], we leverage a GAN-based model to address the labor-intensive data collection problem, which utilizes a small amount of paired Kinect and RFID data to synthesize paired Kinect and RFID data for training the deep learning model. In this paper, *we further improve the work by completely eliminate the need of Kinect data in data augmentation*. Specifically, a forward kinematic layer is introduced to generate synthesized Kinect pose data from simulated human skeleton pose data and quaternions. Then, paired with some pre-collected RFID data, the synthesized Kinect pose data can be leveraged by the GAN to generate synthesized RFID data after training. Therefore, the streamlined generation of synthesized RFID and Kinect data is achieved, and the cost of training data collection for RFID-based pose estimation is effectively reduced. The synthesized Kinect pose data can also be used for other purposes such as the labeled ground truth data for the RFID-Pose model. The high adaptability of the synthesized RFID data is demonstrated by our experiments conducted with and without augmentation.

II. System Design

The RFPose-GAN system is proposed to synthesize high quality RFID data with the smallest cost on training data collection. Fig. 1 presents an overview of the RFPose-GAN system architecture, which is comprised of two key modules: a GAN model and an artificial pose generating layer.

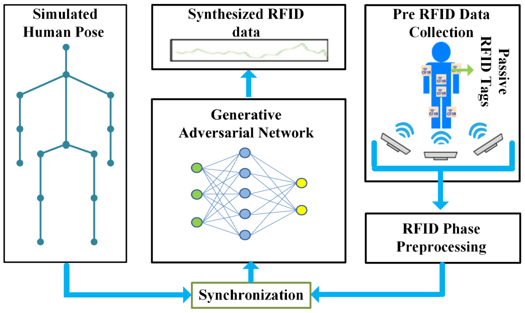

Fig. 1. Overview of the architecture of the proposed RFPose-GAN system.

The human skeleton pose data refers to the 3D coordinates of the subject when performing various activities in front of the Kinect camera. Simulated skeleton poses can be easily reconstructed with a pre-recorded skeleton pose data from [1] to have different limbs and spine lengths, making it possible to create skeleton pose data of people with any body form without using a Kinect camera. An inverse kinematics layer is then used to calculate the unit quaternions for each joint. Since there are only a limited number of rotation combinations for a skeleton in this experiment, more variations are preferred

978-1-6654-5468-1/22/$31.00 ©2022 IEEE

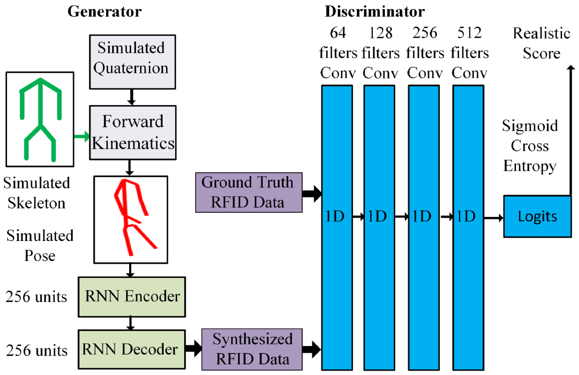

Fig. 2. Illustration of the RFPose-GAN structure.

for greater diversity in the training data. Therefore, we utilize the neural kinematic network architecture in [3] to synthesize unit quaternions for activities such as walking and boxing, and then we introduce white Gaussian noise into the quaternions. With the simulated skeleton pose and quaternions, the forward kinematics layer can hence produce artificial 3D pose data with rich rotation variations.

Fig. 2 presents the proposed GAN model architecture, where an RNN Autoencoder acts as the generator and a 1D convolutional neural network (CNN) serves as the discriminator. Both the Autoencoder and the 1D CNN are chosen for their abilities to extract temporal features. In the generator, the artificial pose data sequences are first fed into the RNN Encoder in which the features are extracted and stored. Then these encoded features are used by the RNN Decoder to synthesize RFID data. Both the RNN Encoder and Decoder utilize 256 gated recurrent units (GRU) as recurrent units. The discriminator then assesses the synthesized RFID data as one of the two inputs. The ground truth RFID data, which is essentially the phase variations gathered from the RFID tags, is the other input. There are a total of 4 hidden layers, with 64, 128, 256, and 512 kernels of widths of 1 in the first, second, third, and fourth hidden layer, respectively. The fourth hidden layer's output is ultimately fed into a final 1D convolutional layer, where it is flattened as a logits vector for accurate score calculation to determine how realistic the synthesized RFID data is; the score helps to further improve the generator.

III. Performance Evaluation

The RFID data is organized as 3rd-order tensors with the third dimension being the three reader antennas throughout 71 consecutive time slots. Fig. 3 shows that for antenna 2, the synthesized RFID data exhibits an overall similarity with the ground truth for each of the joints and throughout time. Note that both the synthesized and ground truth RFID data are smoothed by the Hampel filter for better result showcase.

Five different types of activities are considered in the experiment, including walking, squatting, drinking, handwaving, and boxing. Three datasets are used for training the same RFID-Pose model as in [1], comprised of different combinations of data samples. Dataset_5Act includes the ground truth RFID samples of the five activities. Dataset_Syn consists of only the synthesized samples, while Dataset_Aug has both real and synthesized samples for the five activities. There are only a limited number of samples in this experiment for both real and synthesized samples (i.e., 71 frames of real samples and 213 frames of synthesized samples for each activity), which might hinder the overall performance of the pose reconstruction model, but the efficacy of the synthesized data is still obvious from Fig. 4. The test dataset comprises samples for the same five activity types, but distinct from any of the training datasets. In Fig. 4, the pose estimation errors between the artificial vision poses (i.e., ground truth data) and predicted poses from the RFID data are compared for the three models that are trained with the three datasets respectively. The mean estimation errors for all five activities of the model trained on Dataset_Aug are effectively smaller than those of the model trained on Dataset_5Act, and reasonably smaller than those of the model trained on Dataset_Syn, with the largest error being 7.36cm for boxing. The model trained on Dataset_Syn has one estimation error of 9.27cm for Walking that is larger than that of the model trained on Dataset_5Act, while the rest of the estimation errors are relatively smaller. The experiment results demonstrate that RFPose-GAN can effectively generate useful synthesized RFID data and reduce the data collection cost.

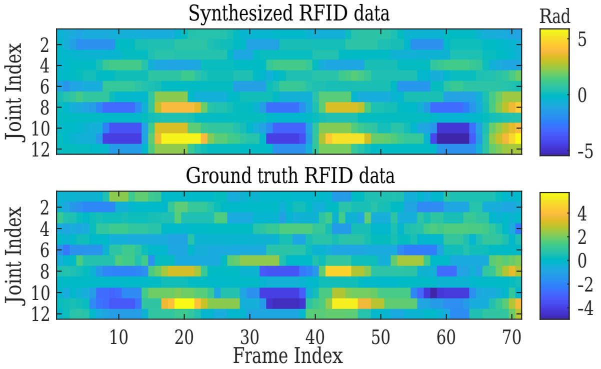

Fig. 3. Example of synthesized RFID data generated by RFPose-GAN.

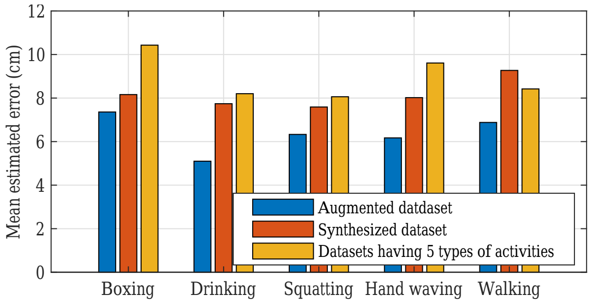

Fig. 4. Mean estimation errors of the three models trained with Dataset_5Act, Dataset_Syn, and Dataset_Aug.

Acknowledgments: This work is supported in part by the NSF under Grants ECCS-1923163 and CNS-2107190

References

[1] C. Yang, X. Wang, and S. Mao, "RFID-Pose: Vision-aided 3D human pose estimation with RFID," *IEEE Transactions on Reliability*, vol. 70, no. 3, pp. 1218–1231, Sept. 2021.

[2] C. Yang, Z. Wang, and S. Mao, "RFPose-GAN: Data augmentation for RFID based 3D human pose tracking," in *Proc. IEEE RFID-TA 2022*, Cagliari, Italy, Sept. 2022, pp.1–4.

[3] R. Villegas, J. Yang, D. Ceylan, and H. Lee, "Neural kinematic networks for unsupervised motion retargetting," in *Proc. IEEE CVPR 2018*, Salt Lake City, UT, June 2018, pp. 8639–8648.