Poster Abstract: 3D Human Pose Estimation Using WiFi Signals

Yili Ren
Florida State University
Tallahassee, FL, USA
ren@cs.fsu.edu

Zi Wang
Florida State University
Tallahassee, FL, USA
ziwang@cs.fsu.edu

Yichao Wang
Florida State University
Tallahassee, FL, USA
ywang@cs.fsu.edu

Sheng Tan
Trinity University
San Antonio, TX, USA
stan@trinity.edu

Yingying Chen
Rutgers University
Piscataway, NJ, USA
yingche@scarletmail.rutgers.edu

Jie Yang
Florida State University
Tallahassee, FL, USA
jie.yang@cs.fsu.edu

ABSTRACT

This paper presents GoPose, a 3D skeleton-based human pose estimation system that uses commodity WiFi devices at home. Our system leverages the WiFi signals reflected off the human body for 3D pose estimation. In contrast to prior systems that need dedicated sensors, our system does not require a user to wear any sensors and can reuse the WiFi devices that already exist in a home environment for mass adoption. To realize such a system, we leverage the 2D AoA estimation of the signals reflected from the human body and the deep learning techniques. Preliminary results show GoPose achieves a high accuracy of 4.5cm in various scenarios.

CCS CONCEPTS

• **Human-centered computing → Systems and tools for interaction design**.

KEYWORDS

WiFi Sensing, Human Pose Estimation, Channel State Information (CSI), Deep Learning

ACM Reference Format:
Yili Ren, Zi Wang, Yichao Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2021. Poster Abstract: 3D Human Pose Estimation Using WiFi Signals. In *The 19th ACM Conference on Embedded Networked Sensor Systems (SenSys '21), November 15–17, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3485730.3492871

1 INTRODUCTION

Estimating the human pose is gaining increasing attention as the human body offers a high degree of freedom for human-computer interactions (HCI). It is a crucial building block to support a variety of emerging applications in smart homes, such as virtual reality, and exercise monitoring. Traditional human pose estimation systems mainly rely on either computer vision techniques or wearable approaches. However, the vision-based systems cannot work in non-line-of-sight (NLoS) scenarios, whereas the wearable systems could be inconvenient. More recently, Radio Frequency (RF) based sensing becomes an appealing alternative for human pose estimation [9] and does not require a user to wear or carry any sensors [6, 8]. It also works under NLoS scenarios [5, 7].

In this work, we propose GoPose, a 3D skeleton-based human pose estimation system by reusing commodity WiFi devices in a home environment. Unlike the prior WiFi-based 3D human pose estimations that only work for a set of predefined activities [2] performed at a fixed position [4], our system works for free-form activities even when the user is moving around, offering on-the-go pose tracking for unseen activities. As GoPose could reuse commodity WiFi devices, it does not incur an additional cost, and thus is promising for mass adoption for end-users in smart homes.

In particular, leverage the two-dimensional (2D) angle of arrival (AoA) of the incident signals derived from the non-linearly spaced antennas to provide spatial information for pinpointing the human body. Then, we utilize the deep learning models of the convolutional neural network (CNN) and the Long Short-Term Memory (LSTM) to abstract the 3D human pose from 2D AoA. In particular, the CNN is used to extract spatial dynamics (e.g., the locations of limbs and the torso), whereas the LSTM is utilized to model temporal dynamics of human poses (e.g., trajectories of limbs and torso).

2 SYSTEM DESIGN

The basic idea of our system is to leverage the spatial information of the 2D AoA and deep learning to model the complex 3D skeletons of the human body for 3D pose estimation. Note that the 3D skeleton consists of multiple joints (i.e., 14 joints listed in Table 1). As illustrated in Figure 1, a WiFi transmitter sends out signals to multiple WiFi receivers to probe human activities. The system takes as input time-series Channel State Information (CSI) measurements. This data is then preprocessed to remove noises. The core of our system is 2D AoA estimation and 3D pose construction from CSI. The system first combines both the spatial diversity and the frequency diversity to increase the resolution of 2D AoA for differentiating signals reflected from different parts of the human body. It then goes through static environment removal to filter out the signals reflected from the indoor environments. After that, the system combines the 2D AoA spectrum of multiple packets at multiple receivers to fully capture the human body.

Next, our system leverages the deep learning models of CNN and LSTM to construct the 3D pose of the human body based on 2D AoA spectrums. CNN captures the spatial feature of the human

SenSys '21, November 15–17, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9097-2/21/11.
https://doi.org/10.1145/3485730.3492871

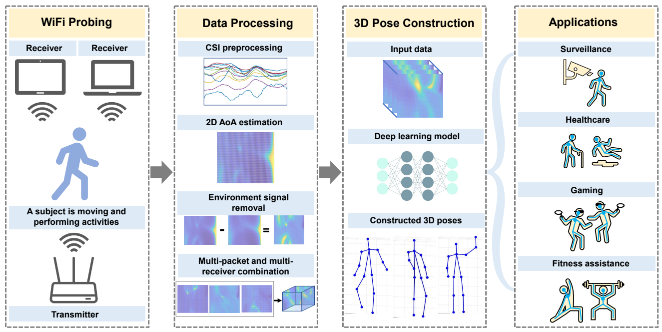

Figure 1: System overview.

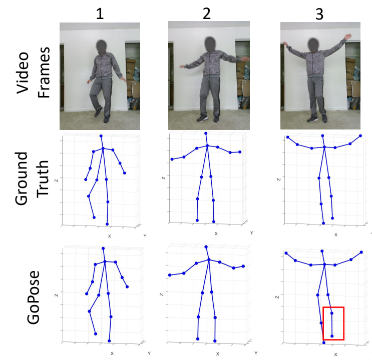

Figure 2: Examples of the constructed 3D skeletons.

Table 1: Average joint estimation errors (unit:*cm*).

Joints	Head	Spine	LShoulder	LElbow	LWrist	RShoulder	RElbow	RWrist	LHip	LKnee	LAnkle	RHip	RKnee	RAnkle	Overall
GoPose	3.3	3.1	4.3	5.7	8.0	4.4	4.9	6.9	3.3	3.4	4.3	3.6	4.0	3.9	4.5

body parts from 2D AoA spectrums, while the LSTM estimates the temporal feature of the motions.

3 PERFORMANCE EVALUATION

We conduct the experiments with five laptops (one transmitter and four receivers). Each laptop is equipped with Intel 5300 NIC connected to three antennas. Linux 802.11 CSI tools [1] are used to extract CSI. We utilize a Microsoft Kinect 2.0 [3] to record the ground truth of the 3D human pose. We evaluate our system in three real-world environments including a living room, a dining room, and a bedroom. Six volunteers (three males and three females) are recruited. Each volunteer is asked to conduct both exergaming activities and everyday activities while she/he is walking around. Our system is trained with 70% of the data set and tested with the remaining 30% of the data set. We use the joint localization error as the evaluation metric. It is defined as the Euclidean distance between the predicted joint location and the ground truth.

Table 1 reports the joint localization error for each joint for GoPose. We can find that the overall localization error is only 4.5cm. The average joint localization error ranges from 3.1cm to 8.0cm. To better visualize the performance of GoPose, we also presented the constructed 3D human skeletons. Figure 2 shows four examples of those constructed 3D skeletons. Although there are a few slight deformations (e.g., in the red solid rectangle), it is easy to observe that majority of the 3D poses estimated by GoPose are highly accurate. These results also demonstrate that the proposed system can accurately construct 3D moving human poses using WiFi signals.

4 CONCLUSION

This paper presents GoPose, a 3D skeleton-based human pose estimation system that offers on-the-go pose tracking in a home environment. In the GoPose system, the 2D AoA spectrum of the signals reflected from the human body is leveraged to locate different parts of the human body, while deep learning is incorporated to model the complex relationship between the 2D AoA spectrums and the 3D skeletons. Preliminary results show that GoPose is highly accurate in constructing 3D human poses.

ACKNOWLEDGMENTS

This research was partially supported by the NSF Grants CNS-2131143, CNS-1910519, and DGE-1565215.

REFERENCES

[1] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review* 41, 1 (2011), 53–53.

[2] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. 2020. Towards 3D human pose construction using wifi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[3] Microsoft. 2021. Kinect 2 for Windows. https://developer.microsoft.com/en-us/windows/kinect/

[4] Yili Ren, Zi Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2021. Winect: 3D Human Pose Tracking for Free-form Activity Using Commodity WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–29.

[5] Sheng Tan and Jie Yang. 2016. WiFinger: Leveraging commodity WiFi for fine-grained finger gesture recognition. In *Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing*. 201–210.

[6] Sheng Tan, Linghan Zhang, Zi Wang, and Jie Yang. 2019. MultiTrack: Multi-user tracking and activity recognition using commodity WiFi. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[7] Sheng Tan, Linghan Zhang, and Jie Yang. 2018. Sensing fruit ripeness using wireless signals. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 1–9.

[8] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. 617–628.

[9] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-based 3D skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. 267–281.