Tracking Free-form Activity 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

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

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

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

ABSTRACT

WiFi human sensing has become increasingly attractive in enabling emerging human-computer interaction applications. The corresponding technique has gradually evolved from the classification of multiple activity types to more fine-grained tracking of 3D human poses. However, existing WiFi-based 3D human pose tracking is limited to a set of predefined activities. In this work, we present Winect, a 3D human pose tracking system for free-form activity using commodity WiFi devices. Our system tracks free-form activity by estimating a 3D skeleton pose that consists of a set of joints of the human body. In particular, Winect first identifies the moving limbs by leveraging the signals reflected off the human body and separates the entangled signals for each limb. Then, our system tracks each limb and constructs a 3D skeleton of the body by modeling the inherent relationship between the movements of the limb and the corresponding joints. Our evaluation results show that Winect achieves centimeter-level accuracy for free-form activity tracking under various environments.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

KEYWORDS

WiFi Sensing, Channel State Information (CSI), Human Pose Estimation, Free-form Activity, 3D Human Skeleton

ACM Reference Format:

Yili Ren, Zi Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. Tracking Free-form Activity Using WiFi Signals. In *The 27th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '21), January 31–February 4, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3447993.3482857

ACM MobiCom '21, January 31–February 4, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-8342-4/22/01...$15.00
https://doi.org/10.1145/3447993.3482857

1 INTRODUCTION

In recent years, WiFi-based sensing is gaining increasing attention due to the prevalence of WiFi devices and their ability to sense the surrounding environments. Indeed, a variety of WiFi-based systems have been proposed to sense various human activities and objects, ranging from large-scale activities [11], indoor localization [6] to small-scale movements [4, 8], daily objects [7, 10], and multi-person tracking [9, 12]. Earlier WiFi human activity sensing mainly focuses on the multi-class classification to distinguish various types of activity and gesture. More recent work has evolved towards constructing a 3D human pose that consists of a set of joints of the body at an unprecedented level of granularity [3]. However, existing WiFi-based 3D human pose tracking is limited to only a set of predefined activities as it relies on the pre-trained model of known activities. It thus cannot work well for free-form activities that were previously unseen by the system. In reality, there exists a variety of emerging Human-Computer Interaction (HCI) applications that demand the 3D human pose of free-form activity. For instance, Virtual Reality applications require capturing the free-form movements of two arms of a player in 3D space. Moreover, medical training in Extended Reality demands free-form motion tracking to enable trainees to learn about surgical operations by using hands and arms to interact with a 3D virtual human body. Additionally, the 3D free-form movement tracking can also enhance the control precision of existing smart home applications, such as continuous and precise thermostat temperature adjustment.

In this work, we propose Winect, a skeleton-based human pose tracking system for free-form activity in 3D space using commodity WiFi devices. Winect does not rely on a set of predefined activities, thus can track free-form movements of multiple limbs simultaneously to enable novel HCI applications. Unlike traditional approaches, Winect does not require cameras or wearable sensors. Winect leverages the WiFi signals reflected off the human body for 3D pose tracking. We propose to model the process of multi-limb signal separation as a blind source separation (BSS) problem. However, solving the BSS problem requires prior knowledge of how many limbs are in motion simultaneously. To address this issue, we develop a limb identification method that leverages the estimation of the two-dimensional (2D) angle of arrival (AoA) of the signals reflected off the human body to infer the number of moving limbs and identify the limbs that are in motion. Then, we can derive the position of each limb in 3D space over time and infer the trajectory of each limb based on the phase changes of the signals. The next challenge is to decompose the trajectory of each limb to the fine-grained trajectories of the joints for 3D pose tracking. To address this issue, we leverage the inherent relationship between the limb and the joints to construct a deep learning model. The evaluation results show that our system can track 3D human poses with centimeter-level accuracy for various free-form activities.

2 SYSTEM DESIGN

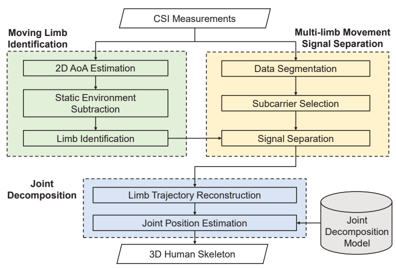

Figure 1: System flow.

The system flow is illustrated in Figure 1, which consists of three major components. Winect first employs *Moving Limb Identification* to identify the number of moving limbs. Next, we conduct *Multi-limb Movement Signal Separation* to separate the multi-limb movement signals. Finally, the obtained positions of limbs go through *Joint Decomposition* and further decompose into joint positions.

2.1 Moving Limb Identification

In this section, we propose a two-dimensional (2D) angle of arrival (AoA) (i.e., azimuth and elevation) estimation method that leverages the L-shaped antenna array, spatial diversity in transmitting antennas and frequency diversity of OFDM subcarriers. These diversities contain enough information that allows our system to jointly estimate 2D AoA. We then identify each moving limb by conducting static environment subtraction and analyzing the signal power change in the azimuth-elevation power spectrum. In order to remove the impact of the static environment, we first derive the 2D AoA spectrum of the static environment without user movement. By subtracting the static environment, we can extract the signal reflected from the moving limb, which is independent of the environment. Next, we detect the peak values in the resulting azimuth-elevation spectrum to identify the number of moving limbs. Here, we conduct the identification process using multiple random CSI packets to mitigate the error. In particular, Winect uses a non-parametric clustering method: Density-based spatial clustering of applications with noise (DBSCAN) algorithm to cluster the peaks without prior knowledge of the peak number. We can easily calculate the average position of each cluster and pinpoint the limb according to the average peak position in the spectrum.

2.2 Multi-limb Movement Signal Separation

After we obtain the number of moving limbs, we can separate movement signals for each limb. The multi-limb movement signal separation can be modeled as a Blind Source Separation (BSS) [13] problem and we can solve the BSS problem by leveraging the Independent Component Analysis (ICA) with the identified number of moving limbs in the previous step. In order to facilitate the signal separation, we need to perform data segmentation. In particular, we divide CSI measurements into a series of 0.1-second segments. The BSS problem can be solved using ICA given that all the sources are independent, non-Gaussian, and combined linearly. Our experiments show that all correlations values are smaller than 0.08. Therefore, we can assume that the movement of different limbs is independent of each other. Our experiments also demonstrate the non-Gaussianity of limb movements. Lastly, we can derive the calibrated CSI: $\mathbf{x}_{calib}(f,t) \approx \frac{\mathbf{s}}{\mu}\mathbf{m}(f,t)$, where $\mathbf{s}$ represents the static environment signals, μ is a complex-valued constant and can be obtained by using the genetic algorithm and $\mathbf{m}(f,t)$ represents the limb movement signals. This shows that the calibrated CSI maintains the linearity of source signals. Thus, we can apply ICA to the calibrated CSI to achieve multi-limb separation. After that, our system will calculate the path length change based on the formulation of limb movement signal $e^{-j2\pi\frac{d_i(t)}{\lambda}}$, where $d_i(t)$ is the path length of the signal reflected from the i^{th} limb, λ is the wavelength. Now, we are able to infer the movement direction and path length change of the limb from a single transmitter-receiver pair. In order to achieve multi-limb tracking in 3D space, we utilize three transmitter-receiver pairs. Given that the distance between each transmitter and receiver is known, we can localize all limb positions and obtain the trajectories by intersecting three ellipsoids with foci at the corresponding transmitter-receiver pair.

2.3 Joint Decomposition

Even after we obtain the free-form trajectory of each limb, it is important to infer the trajectories of the corresponding joints to provide more detailed positions in 3D space. We find that the arm and leg joints have certain limitations for their range of motion and they will also constrain the activities. This inspires us to utilize a deep learning approach to model the inherent relationship between the position of the limb and the corresponding joints. We denote a set of discrete position points (i.e., the position of the limb or the joint) as a point cloud which is shown in Figure 2 and we can easily build the point clouds for both limb positions and joint positions. If the point cloud is dense enough, almost all daily life free-form activities can be represented as the paths formed by points in the point cloud as shown in Figure 3. Thus, we could use the ResNet [2] to learn the relationship between the two point clouds.

3 PERFORMANCE EVALUATION

3.1 Experimental Setup

We conduct experiments with five laptops (i.e., one transmitter, four receivers). Each laptop runs Ubuntu 14.04 LTS and is equipped with Intel 5300 wireless NIC connected with three antennas. Linux 802.11 CSI tools [1] are used to extract CSI data for both limb identification and multi-limb tracking. We set the frequency of the WiFi channel to 5.32 GHz with a bandwidth of 40 MHz. The packet rate is set at 1000 packets per second. In order to obtain the ground truth of the movement of multiple limbs, we utilize a Microsoft Kinect 2.0 [5] and the sampling rate is set at 10 Hz. We conduct experiments

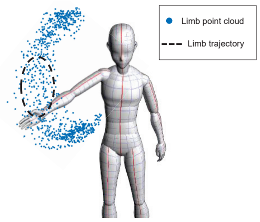

Figure 2: Point cloud for limb positions.

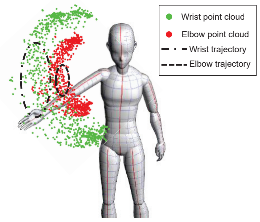

Figure 3: Point cloud for joint positions.

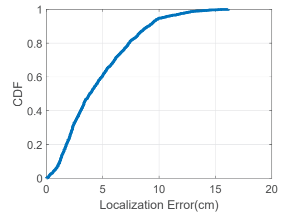

Figure 4: Overall tracking error.

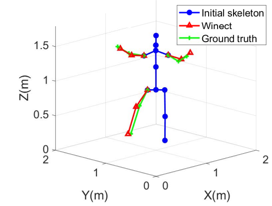

Figure 5: Constructed 3D human skeleton.

Table 1: Joint localization errors (unit:*cm*).

Joints	LElbow	LWrist	RElbow	RWrist	LKnee	LAnkle	RKnee	RAnkle	Overall
Winect	4.2	5.1	4.8	4.9	4.1	4.4	4.3	4.7	4.6

in both the living room and bedroom environments. There are 6 volunteers (3 males and 3 females) who participate in experiments. Each participant is asked to conduct multi-limb free-form daily life movements and there are over 900 activities in total. We utilize the joint localization error, which is defined as the Euclidean distance between the predicted joint location and the ground truth.

3.2 Overall Performance

We evaluate the overall performance of our system with free-from activities. Figure 4 shows the cumulative distribution function (CDF) of tracking errors for our system. We can observe that the median tacking error of our system is 3.9cm. The results demonstrate that our system has high tracking accuracy. Moreover, Table 1 shows the average joint localization error for each joint as well as the overall result for all 8 joints. The range of joint localization error of our system is from 4.1cm to 5.1cm. The overall localization error of our system is 4.6cm. To intuitively observe the performance of Winect, Figure 5 shows an example of the constructed 3D skeleton of free-form activities. We color the initial skeleton with blue, the predicted skeleton with red and the ground truth with green. We can observe that 3D skeletons constructed by Winect are almost the same as the ground truths. The above results demonstrate that our proposed system is able to simultaneously track free-form activities of multiple limbs with high accuracy.

4 CONCLUSION

This paper presents Winect, which is capable of tracking 3D human pose for free-form activities. The proposed system can track free-form motions from multiple limbs simultaneously by reusing existing commodity WiFi devices and does not rely on a set of predefined activities. Our system enables 3D skeleton pose estimation by separating the entangled signals from different limbs and then modeling the relationships between the movements of the limb and the corresponding joints. The experiments show that Winect can track 3D human poses with an average error of 4.6cm for various free-form activities.

ACKNOWLEDGMENTS

We thank the reviewers for their insightful feedback. This work was partially supported by the NSF Grants CNS-1910519, CNS-2131143, and DGE-1565215.

REFERENCES

[1] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review* 41, 1 (2011), 53–53.
[2] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[3] W. Jiang, H. Xue, C. Miao, S. Wang, S. Lin, C. Tian, S. Murali, H. Hu, Z. Sun, and L. Su. 2020. Towards 3D human pose construction using wifi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[4] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng. 2015. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 267–276.
[5] Microsoft. 2021. Kinect 2 for Windows. https://developer.microsoft.com/en-us/windows/kinect/
[6] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu. 2018. Widar2. 0: Passive human tracking with a single wi-fi link. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 350–361.
[7] Y. Ren, S. Tan, L. Zhang, Z. Wang, Z. Wang, and J. Yang. 2020. Liquid Level Sensing Using Commodity WiFi in a Smart Home Environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–30.
[8] S. Tan and J. 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.
[9] S. Tan, L. Zhang, Z. Wang, and J. 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.
[10] S. Tan, L. Zhang, and J. Yang. 2018. Sensing fruit ripeness using wireless signals. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 1–9.
[11] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. 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.
[12] Y. Xie, J. Xiong, M. Li, and K. Jamieson. 2019. mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
[13] Y. Zeng, D. Wu, J. Xiong, J. Liu, Z. Liu, and D. Zhang. 2020. MultiSense: Enabling multi-person respiration sensing with commodity wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–29.