A WiFi Vision-based 3D Human Mesh Reconstruction

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

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

In this work, we present, Wi-Mesh, a WiFi vision-based 3D human mesh construction system. Our system leverages the advances of WiFi to visualize the shape and deformations of the human body for 3D mesh construction. In particular, it estimates the two-dimensional angle of arrival (2D AoA) of the WiFi signal reflections to enable WiFi devices to "see" the physical environment as we humans do. It then extracts only the images of the human body from the physical environment, and leverages deep learning models to digitize the extracted human body into 3D mesh representation. Experimental evaluation under various indoor environments shows that Wi-Mesh achieves an average vertices location error of 2.58cm and joint position error of 2.24cm.

CCS CONCEPTS

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

KEYWORDS

WiFi Sensing, 3D Human Mesh, Deep Learning

ACM Reference Format:
Yichao Wang, Yili Ren, Yingying Chen, and Jie Yang. 2022. A WiFi Vision-based 3D Human Mesh Reconstruction. In *The 28th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '22), October 17–21, 2022, Sydney, NSW, Australia.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3495243.3558247

ACM MobiCom '22, October 17–21, 2022, Sydney, NSW, Australia

ACM ISBN 978-1-4503-9181-8/22/10…$15.00
https://doi.org/10.1145/3495243.3558247

1 INTRODUCTION

Recent years have witnessed tremendous progress in 3D human mesh reconstruction in various applications, such as VR/AR, and virtual try-on [1]. 3D human mesh parameterizes the 3D surface of the human body, which represents how individuals vary in height, weight, somatotype, body proportions, and how the 3D surface deforms with articulation. Traditional approaches primarily rely on computer vision technique, or wearables [1]. These approaches, however, require either significant infrastructure installation or diligent usage of wearable devices. In addition, the computer vision-based systems cannot work well in non-line of sight (NLoS) or poor lighting conditions. They also incur large errors when subjects wear baggy clothes [1].

In this work, we ask whether it's possible to re-use commodity WiFi, originally designed for communication, to construct 3D human mesh. Earlier work has shown that it is able to classify human body activities [7] and detect subtle movements, such as finger gestures [6]. Recently, systems like Winect [3] are proposed to track more detailed 3D human pose for free-form activities. However, none of these systems are able to provide 3D human mesh that consists of thousands of vertices, which is several orders of magnitude larger than the number of body joints defined in 3D poses or activities. In addition, prior systems mainly feed the amplitude/phase or the Doppler frequency shifts of the WiFi signals into deep learning models for activity tracking [5]. They would require more quality training data and deeper neural networks to ensure better system robustness. This limitation renders them less practical across different environments and for unseen people.

In our work, we demonstrate that the commodity WiFi can be leveraged to construct 3D human mesh. We propose a WiFi vision-based approach for 3D human mess reconstruction across different environments and for unseen people. We leverage the advances in WiFi technology to help WiFi devices "see" and visualize the human body as we humans do. Our system, Wi-Mesh, helps WiFi devices "see" a person by leveraging the fairly large number of antennas on the

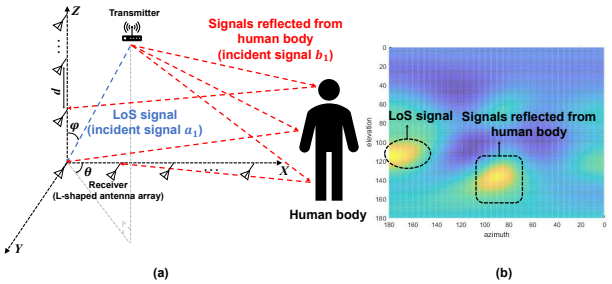

Figure 1: WiFi vision based on 2D AoA.

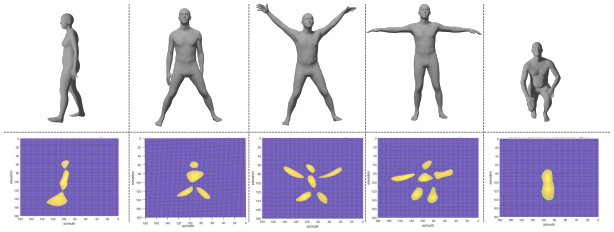

Figure 2: Examples of 2D AoA images.

next-generation WiFi devices. In particular, we estimate two-dimensional angle of arrival (2D AoA) of the signal received at the WiFi receiver in the azimuth-elevation plane, where azimuth is an angular measurement at the horizon direction, and elevation is the angular measurement in the vertical direction. Given the signal intensity of each azimuth-elevation direction, we could derive a visualization or a 2D AoA image of the human body, similar to a gray-scale image captured by a camera.

Given the image of the human body, we design deep learning models to extract both the spatial body shape and temporal body deformations for 3D mesh reconstruction. In particular, both the spatial body shape and temporal body deformations are fitted into the SMPL model [2] to obtain the 3D human mesh, which includes 6890 vertices and 23 joints, as shown in Figure 4.

2 A WIFI VISION-BASED APPROACH

In this work, we propose a concept of 2D AoA-based WiFi vision, which leverage the advances in WiFi technology to help WiFi devices "see" and visualize the physical environment. In particular, the new generation of WiFi 6 or 7 supports up to 8 or 16 antennas. These spatially distributed antennas can be used to separate the signal reflections from different directions/locations, providing spatial information about the physical environment. Thus, the 2D AoA of the WiFi signal reflections can be used to visualize the shape and the poses of the human body.

We illustrate our idea with Figure 1, where an L-shaped antenna array (i.e., N antennas) is used to receive the WiFi signal transmitted from the transmitter. In Figure 1, the L-shaped antenna array is aligned with the X-Z axis. As WiFi signals travel through space, they will be reflected from the human body as well as from other static objects (e.g., walls and furniture) in the environment. We can calculate the 2D AoA of the signal reflections based on the phase shift of the received signals at multiple antennas [3].

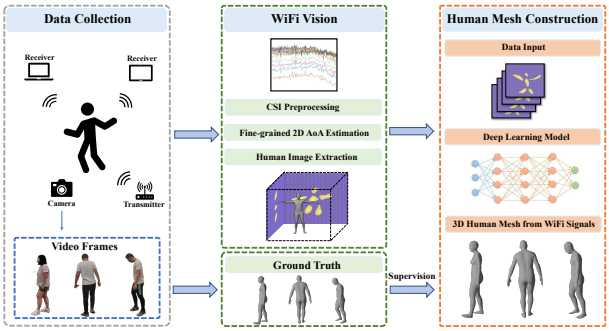

Figure 3: Wi-Mesh system overview.

Figure 1(b) shows an example. The color dots in the figure illustrate the intensity of the reflected signal. We can observe the direction of the LoS and the spatial information of the human body that reflects WiFi signals. This serves as the basis of our proposed WiFi vision-based 3D human mesh construction. To increase the resolution of the 2D AoA image, we further leverage the frequency diversity of the Orthogonal frequency-division multiplexing (OFDM) subcarriers, the spatial diversity of antennas at the transmitter, and the time diversity of WiFi packets. Combining all these diversities provides a better illustration of the shape and deformation of the human body, as shown in Figure 2. The first row shows the ground truth, whereas the second row shows the extracted human body from the 2D AoA spectrums. We can see the silhouettes of the human body and differentiate different poses.

3 SYSTEM DESIGN

As illustrated in Figure 3, our system takes as input time-series CSI measurements at multiple antennas of two WiFi receivers. The WiFi signals reflected from different parts of the human body and surrounding objects will arrive at the receiver in various directions. The CSI measurements then go through reprocess to remove the random phase offsets. Then, our system estimates the 2D AoA of the signals reflected from the human body and static objects by using MUSIC algorithm [4]. Next, our system conducts human image extraction to filter out the signals reflected by the static objects in the environment (e.g., walls and furniture) and only focus on only the human subject. As each 2D AoA image can only capture a subset of the human body due to human body specularity, we further combine multiple 2D AoA images to have a full picture of the human body.

Next, we design a deep learning model to extract both spatial information and temporal deformation of the human body from the 2D AoA images to construct 3D human mesh. The deep learning model has three components: CNN, GRU, and the self-attention module. Among them, CNN is used to parse the static spatial information of the whole human body. The GRU is utilized to extract the dynamic deformation of the body in the temporal dimension. And the self-attention mechanism is used to adaptively learn the contributions of features and highlight the important ones in the final representation. Finally, the extracted spatial body shape and temporal body deformations are fitted into the SMPL model to obtain the 3D mesh representation.

Our system could benefit from the prevalence of WiFi signals and re-use the WiFi devices that already exist in the environment for potential mass adoption. In addition, as the WiFi signals can traverse occlusions/clothes and can illuminate the human body, our system can work under NLoS, poor lighting conditions, or baggy clothes, where the camera-based systems do not work well.

4 EXPERIMENTS

Experimental Setup. We deploy one WiFi transmitter and two WiFi receivers at each experimental site. The transmitter contains three linearly-spaced antennas, and the WiFi receiver has an L-shaped antenna array with two subarrays in the orthogonal direction. The default transmitting packet rate is 1000 packets per second. We capture CSI measurements of 30 OFDM subcarriers for each packet.

Data Collection and Metrics. In the experiments, we recruit 10 volunteers (7 males, 3 females) of different heights and weights to perform random everyday activities. The experiments are conducted in two different real-world environments including a classroom and a living room. And we utilize a camera to record the ground truth for the 3D human mesh. During the training period, we split the data with 10 subjects into two non-overlapping datasets, making sure that the testing data is not seen by the model while training. For evaluation, we leverage the commonly used per vertex error (PVE) and mean per joint position error (MPJPE) for our evaluation.

Overall Performance. In our evaluation, we train our system in the classroom setting, and then test it under the living room environment so as to test whether our system works for different environments and for unseen people. Our experimental results for a single person show that the average PVE and MPJPE of Wi-Mesh are only 2.58cm and 2.24cm, respectively. To qualitatively evaluate the Wi-Mesh, we illustrate the 3D meshes generated by our system for different subjects performing various activities in Figure 4. We use red dotted boxes to highlight the mispredicted and distorted body parts. Nevertheless, we can easily observe that the overall constructed 3D meshes match the ground truth very well.

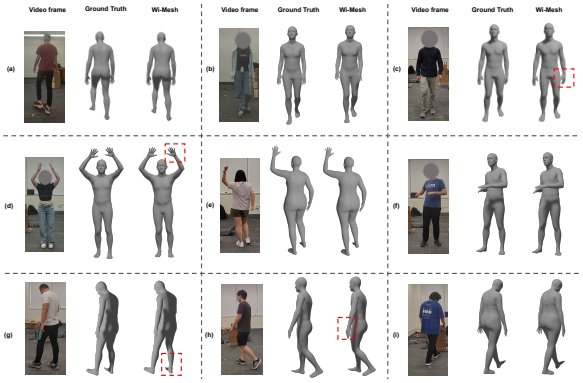

Figure 4: 3D human mesh construction examples.

5 CONCLUSION

In this work, we propose a WiFi vision-based 3D human mesh construction system, which exploits the advances of WiFi and 2D AoA estimation of the signal reflections to visualize the shape and deformations of the human body. Our system then leverages deep learning models to digitize the 2D AoA image of the human body into 3D mesh representation. Experiment results demonstrate that Wi-Mesh is highly effective across different environments and for unseen people.

REFERENCES

[1] Stefano Berretti, Mohamed Daoudi, Pavan Turaga, and Anup Basu. 2018. Representation, analysis, and recognition of 3D humans: A survey. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 14, 1s (2018), 1–36.

[2] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. 2015. SMPL: A skinned multi-person linear model. *ACM transactions on graphics (TOG)* 34, 6 (2015), 1–16.

[3] Yili Ren, Zi Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. Winect: 3D Human Pose Tracking for Free-form Activity Using Commodity WiFi. *ACM ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2022), 1–29.

[4] Yili Ren, Zi Wang, Yichao Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. GoPose: 3D Human Pose Estimation Using WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–25.

[5] Sheng Tan, Yili Ren, Jie Yang, and Yingying Chen. 2022. Commodity WiFi Sensing in 10 Years: Status, Challenges, and Opportunities. *IEEE Internet of Things Journal* (2022).

[6] 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.

[7] 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 *ACM International Conference on Mobile Computing and Networking*. 617–628.