Poster: A WiFi Vision-based Approach to Person Re-identification

Yili Ren
Florida State University
Tallahassee, FL, USA
ren@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
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 propose a WiFi vision-based approach to person re-identification (Re-ID) indoors. Our approach leverages the advances of WiFi to visualize a person and utilizes deep learning to help WiFi devices identify and recognize people. Specifically, we leverage multiple antennas on WiFi devices to estimate the two-dimensional angle of arrival (2D AoA) of the WiFi signal reflections to enable WiFi devices to "see" a person. We then utilize deep learning techniques to extract a 3D mesh representation of a person and extract the body shape and walking patterns for person Re-ID. Our preliminary study shows that our system achieves high overall ranking accuracies. It also works under non-line-of-sight and different person appearance conditions, where the traditional camera vision-based systems do not work well.

CCS CONCEPTS

• **Security and privacy → Security services**.

KEYWORDS

Person Re-identification, WiFi Signals, WiFi Vision

ACM Reference Format:
Yili Ren, Yichao Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. Poster: A WiFi Vision-based Approach to Person Re-identification. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22), November 7–11, 2022, Los Angeles, CA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3548606.3563516

1 INTRODUCTION

Person re-identification (Re-ID) is a fundamental building block for intelligent surveillance. It could enhance public and private security with a wide range of security applications, such as facility security, perimeter monitoring, and abnormal behavior detection. However, traditional camera vision-based systems have several limitations. For example, they cannot work under non-line-of-sight (NLoS) or poor lighting conditions, such as when a person is behind an obstacle or in a dark environment. Moreover, the camera vision-based systems are sensitive to the variation in an individual's appearance due to changes in clothes, viewpoints, and human poses.

Over recent years, WiFi devices have become ubiquitous in public and private places [4, 6, 8]. For example, the number of WiFi reaches 22.2 billion in 2021 and will exceed 31 billion by 2025 [7]. This number is far more than the cameras deployed worldwide. It motivates us to leverage the more pervasive WiFi to illuminate the human body and analyze the reflections for person Re-ID as both visible lights and WiFi signals are electromagnetic waves. Compared to camera vision-based systems, the WiFi-based approach has several advantages, such as working in NLoS or poor lighting conditions. Moreover, the WiFi-based approach is less affected by an individual's appearances such as changes in clothes since the WiFi signal traverses clothes but is reflected off the human body.

Existing work in using WiFi for human sensing uses a black-box approach by directly inputting the received WiFi signals into deep learning models for person identification. Their assumption is that similar activities or people will interrupt the WiFi signals similarly, resulting in similar signal change patterns. This may be useful for applications with pre-defined activities and controlled users but is less applicable for security applications, such as person Re-ID. It is because, given a signal change pattern, it may correspond to a large number of uncontrolled users or many unknown free-from activities. Moreover, the black-box-based deep learning approach is more sensitive to adversarial attacks, where maliciously interrupted WiFi signals can bypass the person Re-ID system [2]. Indeed, recent work has demonstrated such a vulnerability in existing WiFi-based user identification systems [3].

In this work, we propose a WiFi vision-based approach for person Re-ID indoors. We leverage the advancement of WiFi technology and deep learning to help WiFi devices "see", identify, and recognize persons as we humans do. Our system helps WiFi devices "see" a person by leveraging (i) the multiple antennas on WiFi devices, and (ii) the two-dimensional angle of arrival (2D AoA) estimation of the WiFi signal reflections. With spatially distributed antennas, the signal reflections from the different directions could be separated, providing the theoretical foundation to derive spatial information of the physical space. Second, we leverage the 2D AoA (i.e., azimuth and elevation) of the signal reflections to visualize a person. Therefore, the WiFi devices could generate a visualization of the signal reflections from the surrounding environment, thus providing the ability for the WiFi devices to "see" a person. Next, we propose to leverage the deep learning models to digitize the human into a 3D mesh representation. We then extract intrinsic features of a person

CCS '22, November 7–11, 2022, Los Angeles, CA, USA

ACM ISBN 978-1-4503-9450-5/22/11.
https://doi.org/10.1145/3548606.3563516

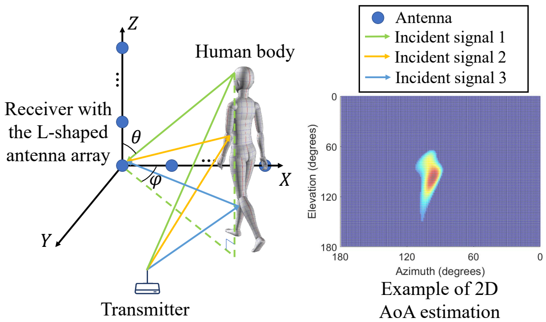


Figure 1: 2D AoA-based WiFi vision.

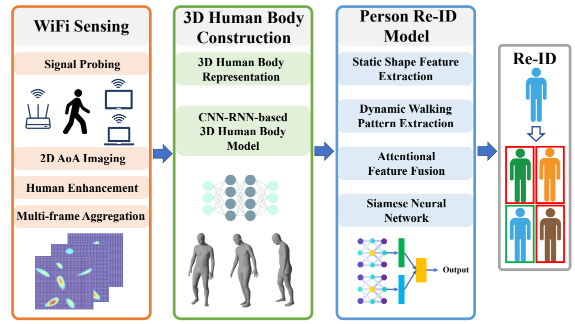
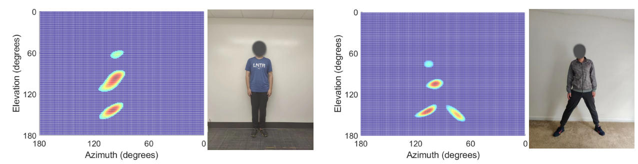


Figure 2: System overview.



(a) A person is standing and the corresponding WiFi vision result. (b) A person is splitting legs and the corresponding WiFi vision result.

Figure 3: Results of WiFi vision.

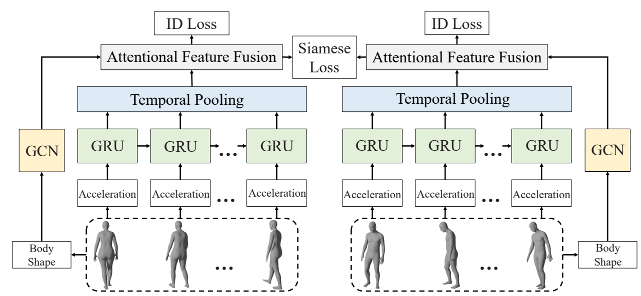


Figure 4: The person Re-ID model.

including both the static body shape and dynamic walking patterns for person Re-ID.

2 SYSTEM DESIGN

2D AoA-based WiFi Vision. We propose to leverage multiple spatially separated antennas to estimate the direction of the signal reflections to visualize a person as the next generation of WiFi supports a large number of antennas at each WiFi device (e.g., WiFi 7 supports up to 16 antennas). Let's assume the multiple antennas on a WiFi receiver form an L-shaped array as shown in Figure 1. Then, the direction of an incident signal to the receiver in physical space (e.g., incident signal 1 in Figure 1) can be uniquely determined by azimuth and elevation (i.e., 2D AoA) [5]. We refer to the estimated 2D AoA spectrum as an "image" or visualization in our work.

The example of the 2D AoA image is shown in the right part of Figure 1, in which the horizontal axis represents the azimuth, the vertical axis represents the elevation and the color represents the signal power. We can observe the shape of a walking person in the spectrum. However, the spatial resolution provided by multiple antennas is still very limited. To improve the resolution of 2D AoA-based images, we utilize four-dimensional information: time diversity of multiple WiFi packets, spatial diversity of both receiving and transmitting antennas, and the frequency diversity of OFDM subcarriers. As shown in Figure 2, our system also performs subject enhancement to segment signal reflections of the human body from irrelevant reflections of the surrounding environments and multi-frame aggregation to yield a more accurate representation of the human body. The results are shown in Figure 3. We can observe that WiFi vision is capable of revealing information about both body shape and pose, thus providing the foundation for WiFi vision-based person Re-ID.

3D Human Mesh Construction. To generate a realistic human mesh, we adopt CNN to extract spatial features from the 2D AoA images. It is because these images contain the general shape and pose information of the human body, CNN can help the system map such information to different vertices of the mesh. We then use a two-layer GRU as the recurrent layer which can model a sequence of temporal dynamics. Because each frame has a distinct effect on the result over time, we use a self-attention technique to dynamically learn the relative contributions of each frame and emphasize the most relevant frames' contributions in the final representation. At last, we map the output of the self-attention layer to the SMPL model to generate the 3D human mesh.

Person Re-ID Model. We build a novel network to extract both static shape and dynamic walking patterns from constructed 3D human meshes for person Re-ID. As shown in Figure 4, we extract static features from the body shape by using a graph convolution network-based method, which considers the 3D mesh as a graph and produces a vector representing the person's static shape biometrics. The dynamic walking pattern features contain the gait pattern, arm and torso gestures, head and body movements. Thus, we compute the acceleration of each joint of the 3D human mesh. Then, we make acceleration information flow between time steps by using the GRU and all time steps are combined using temporal pooling. Next, these static shape and dynamic walking pattern features are fused together using an attention-based mechanism. Lastly, the two-stream sub-networks for two 3D mesh sequences from two different people are constructed following the Siamese network architecture, in which the parameters of sub-networks are shared.

Given a pair of 3D mesh sequences from the same person, the Siamese architecture is trained to produce feature vectors that are close in feature space, while given a pair of 3D mesh sequences from different persons, the network is trained to produce feature vectors that are separated. Given the feature vectors (F_i, F_j) for person i and person j, we utilize the Euclidean distance Hinge loss $L_{Siamese}(F_i, F_j)$ to train our model. We also calculate the identity loss $L_{ID}(F_i)$ and $L_{ID}(F_j)$ using the cross-entropy loss. The final loss function is written as $L_{ReID} = L_{Siamese} + L_{ID}(F_i) + L_{ID}(F_j)$.

3 PERFORMANCE EVALUATION

Devices. We conduct experiments using one WiFi transmitter and two receivers. The WiFi transmitter is equipped with a linear antenna array of three antennas. Each receiver is equipped with an L-shaped antenna array of nine antennas. Linux 802.11 CSI tools [1] are used to extract CSI measurements. We utilize a camera to record the ground truth for both the 3D mesh and Re-ID for the person.

Dataset. In this work, 12 participants (8 males, 4 females) were recruited for our experiments. We collected synchronized WiFi and vision data of random everyday activities of these participants over two months time period. Moreover, we ask these participants to walk randomly for person Re-ID. For calculating ranking accuracy, the whole dataset is randomly split into two non-overlapping parts: 50% of people for training and the remaining 50% of people for testing. The experiments are repeated 10 times with different training and test splits and the results are averaged to ensure stable results.

Evaluation Metrics. We leverage the ranking accuracy to evaluate our system, which is a common evaluation metric for person Re-ID. The system is given a WiFi sample of a test person and only one of the candidates can match the queried WiFi sample of the person. The top-k accuracy is defined as the percentage of cases where the correct test person is ranked among the top k positions of all the candidates in a test. We also show the performance with the average Cumulative Matching Characteristics (CMC) curves.

Overall Performance. We study the overall performance of our system under both line-of-sight (LoS) and NLoS scenarios. For both scenarios, the person can appear in one indoor environment and walk into another one or appear in the same environment at different times with different walking trajectories or appearances. As shown in Figure 5, our proposed system achieves good performance in the LoS scenario. In particular, our system has top-1, top-2 and top-3 accuracies of 85.7%, 92.5%, 97.4%, respectively. This shows our system achieves high overall ranking accuracies as our system focuses on the 3D body shape and walking style of a person, which remain unaffected even when the person wears different clothes at different times and localizations.

We also evaluate the impact of NLoS by placing a wooden screen between the person and each pair of transceivers. We note that the model is trained under the LoS scenario and tested in the NLoS scenario. We can observe that our system still achieves a rank-1 accuracy of 83% and rank-3 accuracy of 95%. Such a result shows that our system can identify people even under the NLoS scenario, where the computer vision-based systems will fail completely. This is because the WiFi signals can penetrate obstacles even if they may be attenuated, while visible light will be totally blocked. As a result, we can deploy the proposed system under more real-world scenarios that do not have LoS or under poor lighting conditions.

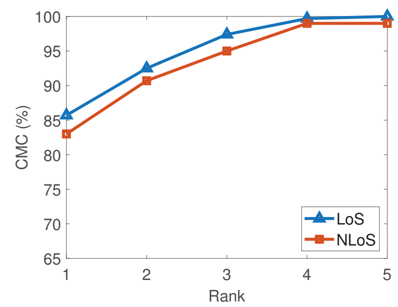


Figure 5: Overall system performance.

4 CONCLUSION

Person Re-ID in traditional optical camera-based systems is challenging due to changes in the appearance of people, occlusions, and unconstrained human poses. We propose a WiFi vision-based person Re-ID system which is very promising to mitigate these challenges. Specifically, we exploit multiple antennas on WiFi devices and 2D AoA of the WiFi signal reflections to visualize a person. Our system extracts intrinsic features of the body shape and dynamic walking patterns from the digitized 3D human mesh for person Re-ID. Our system is thus resistant to the changes in the appearance of people as well as the unconstrained poses. Experiments demonstrate that our system is effective in identifying a number of people and that it can achieve high overall accuracy.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful feedback. This work was partially supported by the NSF Grant CNS-1910519.

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] Chi Lin, Jiaye Hu, Yu Sun, Fenglong Ma, Lei Wang, and Guowei Wu. 2018. WiAU: An accurate device-free authentication system with ResNet. In *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 1–9.

[3] Jianwei Liu, Chaowei Xiao, Kaiyan Cui, Jinsong Han, Xian Xu, and Kui Ren. 2022. Behavior Privacy Preserving in RF Sensing. *IEEE Transactions on Dependable and Secure Computing* (2022).

[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] 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.

[6] 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).

[7] Lionel Sujay Vailshery. 2021. IoT and non-IoT connections worldwide 2010-2025. https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/.

[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.