Poster Abstract: A Vision-based Approach for Commodity WiFi Sensing

Yili Ren*, Yichao Wang*, Yingying Chen§, Jie Yang*
Florida State University* Rutgers University§
{ren, ywang, jie.yang}@cs.fsu.edu yingche@scarletmail.rutgers.edu

ABSTRACT

The ubiquitous WiFi signals provide us the opportunity to sense human activities and the physical environment. In this work, we take a layered approach to design a vision-based method for commodity WiFi sensing. Specifically, the next-generation WiFi supports a larger number of antennas that can provide spatial information of the signal reflections, which enables a vision-based approach for WiFi sensing. To better leverage the spatial formation of the signal reflections and fulfill emerging applications, we provide a holistic layered framework including hardware, physical, deep learning, and application layers as well as a case study. The proposed layered approach could enlighten the research on future WiFi sensing.

CCS CONCEPTS

• **Human-centered computing → Visualization systems and tools**.

KEYWORDS

Channel State Information (CSI), WiFi Sensing, WiFi Vision

ACM Reference Format:
Yili Ren*, Yichao Wang*, Yingying Chen§, Jie Yang* . 2022. Poster Abstract: A Vision-based Approach for Commodity WiFi Sensing. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22), November 6–9, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3560905.3568068

1 INTRODUCTION

WiFi sensing has gained increasing popularity and enabled many emerging applications. The ubiquitous WiFi signals and devices in the Internet of things (IoT) era provide us the opportunity to extend the capabilities of WiFi beyond communication. WiFi sensing can be divided into several generations based on the development of WiFi sensing techniques. The received signal strength (RSS)-based approach is one of the early approaches, especially for device-based localization [7, 9]. Then Channel State Information (CSI) attracts many research efforts as it provides much more fine-grained information than RSS [4]. Thus, many CSI-based sensing systems emerge for various applications with the help of deep learning [5]. However, directly using CSI is less robust or explainable to more complicated applications such as multi-user sensing, human pose/mesh estimation, or complex scene sensing since they offer no explicit spatial information for people, objects, or environments [3].

The next-generation WiFi is capable of supporting a large number of antennas, which can be utilized to discern the signal reflected from spatially separated objects and people. Thus, the next-generation WiFi could become the driving force behind a new generation of WiFi sensing, which provides spatial resolution similar to that of the optical image. The most recent WiFi sensing studies [2] propose to extract a two-dimensional (2D) angle of arrival (AoA) (i.e., azimuth and elevation) of the WiFi signal reflections, which could visualize the human activities or physical objects in the physical space. Moreover, we can further extend 2D AoA information to higher dimensional information by jointly estimating the time of flight (ToF) and Doppler frequency shift (DFS) [8]. In contrast to many existing WiFi sensing systems that use a black-box approach by directly inputting the received WiFi signals into deep learning networks, the 2D AoA-based signal reflections provide domain knowledge of the human activities and physical objects to deep learning. We mainly refer to the 2D AoA-based visualization approach as WiFi vision, which could be the driving force behind a new generation of WiFi sensing.

2 THE LAYERED FRAMEWORK

To leverage WiFi vision and fulfill various applications, we essentially provide a framework with four layers: hardware, physical, deep learning, and application. First, we introduce the WiFi hardware that enables WiFi vision-based sensing. We then propose the data model for the WiFi physical layer. Next, we summarize the design of the deep learning network for segmentation, feature extraction, etc. Finally, we discuss the important applications enabled by the WiFi vision-based commodity WiFi sensing.

Hardware Layer. This layer forms the cornerstone of the approach as shown in Figure 1. Previous generations of WiFi such as WiFi 4 and WiFi 5 have been widely utilized for WiFi sensing but only have a limited number of antennas (e.g., 2 or 3). The new generation of standards in WiFi technology (e.g., WiFi 6 and WiFi 7) can support a large of number antennas (e.g., 16×16 MIMO) and larger bandwidth (e.g., 320MHz bandwidth). Such an advancement along with MIMO technologies enables us to distinguish a larger number of signal paths using the antenna array on a single device. This implies that more and more spatial information in WiFi signals could be obtained as AoA and ToF resolutions are dependent on antenna numbers and bandwidth, respectively.

Physical Layer. Existing CSI tools [1] can extract CSI measurements from COTS WiFi NICs. Prior to conducting WiFi vision using 2D AoA, CSI denoising is required to remove CSI phase noises. It is because the WiFi device's hardware imperfections could lead to

SenSys '22, November 6–9, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9886-2/22/11…$15.00
https://doi.org/10.1145/3560905.3568068

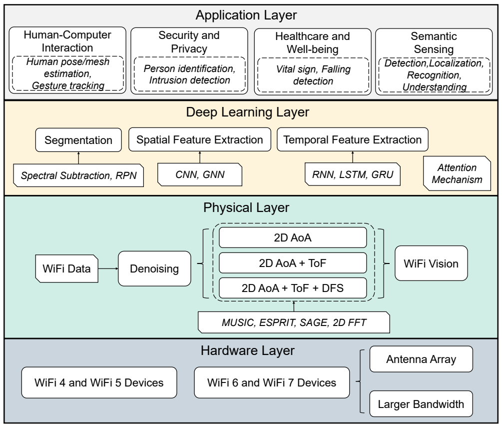


Figure 1: A layered framework for WiFi vision-based sensing.

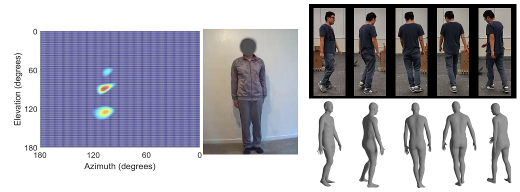
CSI phase distortions. We propose to realize WiFi vision using 2D AoA in terms of the azimuth and elevation angles. Azimuth is an angular measurement of the signal on the horizon and elevation is the angle at which the same signal is measured in the vertical direction. Several algorithms have been advocated for efficient and accurate joint estimation of the 2D AoA such as MUSIC [2].

Deep Learning Layer. The WiFi vision approach could capture the whole environment including the static environment/background and multiple dynamic targets/people. Therefore, the goal of segmentation is twofold: removing the environment and separating multiple targets. They can be done by spectral subtraction and region proposal networks, respectively. The WiFi vision results are image data and contain plenty of spatial-temporal features. Thus, many deep learning techniques (e.g., CNN, RNN, etc.) can be reused for feature extraction. Moreover, the performance of the deep learning networks for WiFi sensing can be improved by the attention mechanism, which enhances the exploration of the relationship between the input WiFi data and the inference outcome.

Application Layer. Many emerging IoT applications may be enabled by WiFi vision-based sensing. For example, by taking the WiFi vision results of a human as input, the deep learning network can predict the 3D human pose/mesh for human-computer interaction. WiFi vision approach may provide us an opportunity to achieve intelligent surveillance systems under non-line-of-sight (NLoS) and could augment traditional camera-based systems. Also, it may assist caregivers in continuously monitoring patients and reporting any urgent situations. WiFi vision can also be a promising approach to semantic sensing including detection, localization, recognition, and understanding of human activities.

3 CASE STUDY

We illustrate the proposed layered framework for WiFi vision with an application of 3D human mesh estimation [6]. For hardware, we use one WiFi transmitter and two receivers. The transmitter is equipped with a linear antenna array and each receiver has an L-shaped array of nine antennas. The WiFi channel is set at 5.32 GHz with 40 MHz bandwidth. In the physical layer, we take the denoised WiFi CSI data as input of the MUSIC algorithm to achieve 2D AoA-based WiFi vision as shown in Figure 2(a). Next, the WiFi vision results go through the deep learning layer to fulfill the application. In particular, we separate the subject from the



(a) An example of 2D AoA-based WiFi vision. (b) Examples of 3D human mesh estimation using the WiFi vision-based approach.

Figure 2: Case study.

background environment. We then leverage CNNs and GRUs to extract both spatial and temporal features of the subject. As last, the features are fused via the attention mechanism. We further map the fused features to the 3D human meshes. The results are shown in Figure 2(b), in which the first row shows the ground truth RGB images captured by cameras, and the second row shows 3D human meshes estimated by our system. The average per vertex error (PVE) of our system is only 2.8cm. We can also observe that the majority of the 3D meshes estimated by our system are highly accurate. This demonstrates that our proposed layered approach to WiFi vision is effective and could support emerging applications. We note that our system can work in non-line-of-sight (NLoS) conditions as WiFi signals can penetrate obstacles.

4 CONCLUSION

In this work, we utilize a layered approach to reflect the hierarchical designs for WiFi sensing based on WiFi vision. The implicit spatial information in WiFi vision could be the driving force behind a new generation of WiFi sensing. The layered framework includes hardware, physical, deep learning, and application layers and we discuss the role and functionality of each layer. We offer a summary of this emerging research area and thus motivate researchers to develop a new generation of WiFi sensing technologies.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful feedback. This work was partially supported by the NSF Grant CNS-2131143 and CNS-1910519.

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] Y. Ren, Z. Wang, Y. Wang, S. Tan, Y. Chen, and J. 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.
[3] S. Tan, Y. Ren, J. Yang, and Y. Chen. 2022. Commodity WiFi Sensing in 10 Years: Status, Challenges, and Opportunities. *IEEE Internet of Things Journal* (2022).
[4] 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.
[5] 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.
[6] Y. Wang, Y. Ren, Y. Chen, and J. Yang. 2022. Wi-Mesh: A WiFi Vision-based Approach for 3D Human Mesh Construction. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*.
[7] Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin. 2013. Measuring human queues using WiFi signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*. 235–238.
[8] 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.
[9] J. Yang and Y. Chen. 2009. Indoor localization using improved rss-based lateration methods. In *GLOBECOM 2009 IEEE Global Telecommunications Conference*. IEEE, 1–6.