WiFi CSI-based Long-Range Person Localization Using Directional Antennas

Julian Strohmayer & Martin Kampel
Computer Vision Lab, TU Wien
Favoritenstr. 9/193-1, 1040 Vienna, Austria
{julian.strohmayer,martin.kampel}@tuwien.ac.at

Abstract

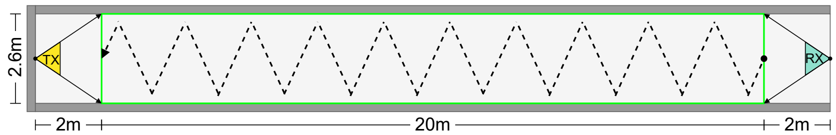
To address limited sensing hardware options in WiFi-based Human Activity Recognition (HAR), we introduce a compact, cost-effective system pairing the *Espressif ESP32-S3* microcontroller with a directional antenna for long-range sensing. Constructed exclusively with widely available off-the-shelf components, our solution is highly reproducible. Moreover, its ability to capture person-centric information over extended distances is evaluated in a localization experiment covering an area of 2.6m×20m. In this experiment, a 3D location regression model, trained on Channel State Information (CSI) amplitude spectrograms and the walking trajectories of a person, achieves a promising test RMSE of 0.197m.

1 Introduction and Related Work

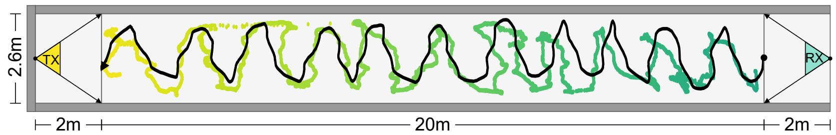
Within the field of HAR, while optical modalities are prevalent, WiFi is gaining recognition due to advantages such as cost-effectiveness, unobtrusiveness, immunity to illumination changes, and visual privacy protection Fu et al. (2020); Arning & Ziefle (2015). Its ability to penetrate walls enables long-range sensing in confined indoor environments, offering economic benefits and enabling innovative applications like through-wall HAR Strohmayer & Kampel (2023). Despite the ubiquity of WiFi devices, CSI capture is limited to specific hardware and software combinations Schumann et al. (2023). Solutions include NICs like *Intel Wireless Link 5300* with *Linux 802.11n CSI Tool* Halperin et al. (2011), *Atheros* NICs with *Atheros CSI Tool* Xie et al. (2015), newer platforms like *Raspberry Pi* and smartphones using *Nexmon CSI Tool* Gringoli et al. (2019), and routers like *Asus RT-AC86U* utilizing *AX-CSI Tool* Gringoli et al. (2021). Finally, CSI capture is now also possible on *ESP32* microcontrollers via *Espressif's IoT Development Framework* (ESP-IDF)[1] and *Wi-ESP* Atif et al. (2020), offering cost advantages and stand-alone CSI sensing capability. Like with the *Raspberry Pi* and smartphones, a potential drawback, however, is the limitation to a single antenna. In WiFi-based HAR, the *ESP32* has established itself as a viable development platform, commonly leveraging the built-in printed inverted F-antenna (PIFA)[4] Schumann et al. (2023). However, we argue against the PIFA's suitability for long-range HAR due to its omnidirectionality and low gain of 2 dBi, resulting in an inability to constrain the recording environment and susceptibility to noise. These shortcomings can be addressed with directional antennas, as explored in this work, marking the second instance of combining the *ESP32* with a directional antenna for WiFi-based HAR Strohmayer & Kampel (2023). In comparison, our system improves reproducibility by only using commercial off-the-shelf components, achieves similar sensing performance in a smaller form factor, and incorporates a powerful single-board computer for real-time on-device inference.

2 Proposed System

The proposed WiFi system, illustrated in Figure 1c, integrates CSI sensing and processing hardware within a compact 3D-printed enclosure. At its core is the *Espressif ESP32-S3-DevKitC-1U*[2], featuring the *ESP32-S3-WROOM-1U*[3] microcontroller for WiFi connectivity and CSI access via ESP-IDF[1]. The *ESP32-S3-WROOM-1U* includes an I-PEX MHF1 connector, enabling the use of an external antenna and bypassing the PIFA present in most *ESP32* variants. Our system employs the *ALFA Network APA-M25*[5], a USD 20 dual-band directional panel antenna with a 66° horizontal beam width and 8dBi gain @2.4GHz. It shares characteristics with the antenna used in Strohmayer & Kampel (2023) (70° beam width and 10-12dBi gain) but is more robust and reproducible as a commercial product. As a third (optional) component, the *Nvidia Jetson Orin Nano*[6] is utilized for CSI packet recording and real-time on-device inference. The proposed system can be configured with different antennas, such as the *ESP32* PIFA (by switching to the *Espressif ESP32-S3-DevKitC-1*[3]), or by connecting different external antennas to the I-PEX MHF1 connector of the *Espressif ESP32-S3-DevKitC-1U*. System CAD models and a bill of materials are made publicly available[9].



(a) Point-to-point transmitter-receiver arrangement in the evaluation environment.



(b) Test trajectory (black) and predicted locations (colored).

(c) WiFi system.

Figure 1: (a) Experimental setup showing the point-to-point transmitter-receiver arrangement in the evaluation environment, with the recording area highlighted in green and the approximate shape of walking trajectories (dotted line). (b) Walking trajectory of the test sequence (black) and locations predicted by our model (colored). (c) Proposed WiFi system used for CSI transceiving.

3 Localization Experiment

To demonstrate the system's capability of capturing person-centric information over large distances, an indoor person localization experiment is conducted as described in the following. **Environment.** Figure 1a shows our experimental setup in a hallway environment. Identical transmitter and receiver systems, connected via *ESP-NOW*[7] at a 100Hz packet sending rate, are arranged in a point-to-point configuration with a 24m spacing. To ensure full horizontal beam coverage of the recording area (highlighted in green), a 2m offset on both ends yields a size of 2.6m$\times$20m. **Data Collection.** For supervised training of a regression model that can predict a person's 3D location in the recording area from CSI, we create the HAllway LOCalization (HALOC) dataset. This dataset comprises six walking sequences, each 4-5 minutes long (4 training, 1 validation, and 1 test). Sequences are acquired by jointly capturing CSI packets and egocentric video (chest-mounted camera) while a person walks from the receiver to the transmitter, as shown in Figure 1a (dotted line). Ground truth locations are extracted from egocentric videos using *ORB-SLAM3* Campos et al. (2021), resulting in location time series sampled at 30Hz. To match the CSI sampling rate of 100Hz, location time series are linearly up-sampled. The HALOC dataset is made publicly available to facilitate further research[9]. **Model Training.** Utilizing the HALOC dataset, we train a 3D location regression model based on the *EfficientNetV2 small* architecture Tan & Le (2021) implemented by *torchvision.models*[8]. The model takes CSI amplitude spectrograms as input, constructed from the amplitudes of 52 L-LTF subcarriers over a fixed number of WiFi packets. After a hyperparameter search for the spectrogram width $w \in \{11, 21, ..., 501\}$, the optimal value is identified as $w = 351$ (approximately 3.5 seconds at a 100Hz sampling rate), resulting in an input spectrogram size of 52$\times$351. The training hyperparameters and data augmentations are given in Appendix B. **Results.** Quantifying localization performance with Root Mean Squared Error (RMSE), our model achieves an error of 0.197m on the HALOC test sequence. This result is visualized in Figure 1b, showing that the model is able to reconstruct the overall shape of the test trajectory, capturing directional changes. This demonstrates the feasibility of long-range person localization with the proposed system, providing a potentially low-cost, easy-to-deploy platform for CSI-based HAR applications. Furthermore, through the integration of the *Nvidia Jetson Orin Nano*, the proposed system enables real-time applications. When deployed, our regression model achieves an inference time of 68.36$\pm$0.34ms (14.63 fps).

4 Limitations

While the *EfficientNetV2 small* regression model achieves promising localization performance on the HALOC dataset, it merely serves as a reproducible proxy methodology to demonstrate that the proposed system is capable of capturing person-centric information over long distances. Achieving generalization across diverse environments in practice, an open problem in WiFi-based HAR will require additional techniques Chen et al. (2023), not addressed in this work.

5 Conclusion

In this work, we proposed a novel WiFi system that combines the *ESP32-S3* with a directional antenna for long-range HAR applications based on WiFi CSI. To evaluate the system's capability for person localization, we deployed it in a hallway environment to collect the HALOC dataset, which comprises WiFi CSI time series and walking trajectories. Trained on this data, a regression model based on the *EfficientNetV2 small* architecture achieved a promising test RMSE of 0.197m.

URM Statement

The authors acknowledge that at least one of the authors of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

References

Katrin Arning and Martina Ziefle. "get that camera out of my house!" conjoint measurement of preferences for video-based healthcare monitoring systems in private and public places. In Antoine Geissbühler, Jacques Demongeot, Mounir Mokhtari, Bessam Abdulrazak, and Hamdi Aloulou (eds.), *Inclusive Smart Cities and e-Health*, pp. 152–164, Cham, 2015. Springer International Publishing. ISBN 978-3-319-19312-0.

Muhammad Atif, Shapna Muralidharan, Heedong Ko, and Byounghyun Yoo. Wi-ESP—A tool for CSI-based Device-Free Wi-Fi Sensing (DFWS). *Journal of Computational Design and Engineering*, 7(5):644–656, 05 2020. ISSN 2288-5048. doi: 10.1093/jcde/qwaa048.

Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021.

Chen Chen, Gang Zhou, and Youfang Lin. Cross-domain wifi sensing with channel state information: A survey. *ACM Computing Surveys*, 55(11):1–37, 2023.

Biying Fu, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Sensing technology for human activity recognition: A comprehensive survey. *IEEE Access*, PP:1–1, 01 2020. doi: 10.1109/ACCESS.2020.2991891.

Francesco Gringoli, Matthias Schulz, Jakob Link, and Matthias Hollick. Free your csi: A channel state information extraction platform for modern wi-fi chipsets. In *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, WiNTECH '19, pp. 21–28, 2019. URL `https://doi.org/10.1145/3349623.3355477`.

Francesco Gringoli, Marco Cominelli, Alejandro Blanco, and Joerg Widmer. Ax-csi: Enabling csi extraction on commercial 802.11ax wi-fi platforms. WiNTECH '21, pp. 46–53, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450387033. doi: 10.1145/3477086.3480833. URL `https://doi.org/10.1145/3477086.3480833`.

Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM CCR*, 41(1):53, Jan. 2011.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Robert Schumann, Frédéric Li, and Marcin Grzegorzek. Wifi sensing with single-antenna devices for ambient assisted living. pp. 1–8, 10 2023. doi: 10.1145/3615834.3615841.

Julian Strohmayer and Martin Kampel. Wifi csi-based long-range through-wall human activity recognition with the esp32. In *Computer Vision Systems*, pp. 41–50, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-44137-0.

Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pp. 10096–10106. PMLR, 2021.

Yaxiong Xie, Zhenjiang Li, and Mo Li. Precise power delay profiling with commodity wifi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, pp. 53–64, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3619-2. doi: 10.1145/2789168.2790124. URL `http://doi.acm.org/10.1145/2789168.2790124`.

A Appendix

Table 1: Online resources (accessed November 22, 2023).

Description	URL
[1]ESP-IDF	https://docs.espressif.com/projects/esp-idf/en/latest/esp32/get-started/
[2]ESP32-S3-DevKitC-1 (U)	https://docs.espressif.com/projects/esp-idf/en/latest/esp32s3/hw-reference/esp32s3/user-guide-devkitc-1.html
[3]ESP32-S3-WROOM-1U	https://www.espressif.com/sites/default/files/documentation/esp32-s3-wroom-1_wroom-1u_datasheet_en.pdf
[4]ESP32 PIFA	https://www.ti.com
[5]ALFA Network APA-M25	https://alfa-network.eu/apa-m25
[6]Nvidia Jetson Orin Nano	https://developer.nvidia.com/buy-jetson?product=all&location=US
[7]ESP-NOW	https://www.espressif.com/en/solutions/low-power-solutions/esp-now
[8]PyTorch *torchvision.models*	https://pytorch.org/vision/stable/models/efficientnetv2.html
[9]Supplementary Material	https://zenodo.org/records/10715595

B Training Details

For model training, we optimize for Mean Squared Error (MSE) loss with AdamW Loshchilov & Hutter (2017) at a learning rate and weight decay of 0.001. Additionally, a cosine annealing learning rate scheduler Loshchilov & Hutter (2016) is employed. To augment training spectrograms, random channel-wise amplitude perturbations (± 0.2), pixel-wise dropout ($p = 0.2$), and column-wise dropout ($p = 0.2$) using the channel mean as a replacement, are applied. Training is conducted with a batch size of 16 for 200 epochs, and the model instance with the best validation performance is selected for evaluation on the test sequence.

C Acknowledgements

This work is partially funded by the Vienna Business Agency (Blindsight-grant 4829418).