

Application of Wireless Sensing in Human Pose Estimation

Junxi Huang

School of Telecommunications Engineering, Xidian University, Xi'an, Shanxi, 710071, China

24012100067@stu.xidian.edu.cn

Abstract. Human pose estimation is widely used in various fields, including smart home, medical treatment, security monitoring, and virtual reality. Traditional human pose estimation methods based on optical images are confronted with issues such as object occlusion and privacy protection. In contrast, wireless sensing technics have attracted attention due to their excellent privacy protection and adaptability to various circumstances. Taking Wi-Fi signals and millimeter-wave radar as examples, this paper deeply explores their current application status and representative works in human pose estimation. In this research, it is found that Wi-Fi signals demonstrate great potential in scenarios such as smart home interactive control and health monitoring. By virtue of high resolution and penetrability, millimeter-wave radar shows obvious advantages in scenarios such as security monitoring, industrial and medical applications. In the future, the development of wireless sensing in human pose estimation can be further promoted by means such as 6G technology, algorithm optimization, and federated learning to enhance privacy protection, providing technical support for multiple fields.

INTRODUCTION

Human Pose Estimation acts as a critical technology in applications such as intelligent interaction, behavior analysis, and health monitoring. Traditional methods such as cameras and voice recognition have problems like object occlusion and privacy leaks. Wireless sensing technology, which utilizes wireless signals such as Wi-Fi and millimeter-wave radar for non-contact sensing, has become an emerging research direction in the field of human pose estimation due to its strong penetration and good privacy protection. This method is expected to overcome the limitations of traditional methods and promote the further development of related applications.

In terms of Wi-Fi signals, Widar 3.0, the cross-domain gesture recognition system proposed by Yi Zhang et al. [1], utilized the channel state information (CSI), extracted domain-independent characteristic body-coordinate velocity profile (BVP), constructed a universal, and finally achieved the zero-effect cross-domain gesture recognition. Based on this system, Liang Fang et al. [2] further optimized the local deployment of this system in the realm of smart homes, reducing costs and enhancing its practical application value. Xiang Zhang et al. [3] proposed an objective gesture recognition system GESFI, which employs a method combining self-improvised pseudo-labeling strategy with adversarial learning. Through mining latent domain labels and learning domain-invariant features, it addresses the issue of cross-domain data distribution differences and enhances the model's generalization ability in multi-domain scenarios. Kangwei Yan et al. [4] proposed an end-to-end multi-person 3D pose estimation system, Person-in-WiFi 3D. This system utilized multi-receiver Wi-Fi signals in combination with a Transformer frame to achieve end-to-end multi-person 3D pose estimation, reaching comparable accuracy to that of cameras and millimeter-wave radars on a self-built dataset. In terms of millimeter-wave radar, a human action recognition method for millimeter-wave radar based on a Dynamic GNN-MB network was proposed by Guoliang Peng et al. [5]. This method adaptively learnt the edge weights of sparse point clouds through a dynamic edge selection function and combined stacked BIGRU to capture temporal features, effectively addressing the problems of voxelization dependence and feature redundancy in traditional methods, incorporating the advantages of privacy protection and high accuracy. Haoze Du et al. [6] proposed a millimeter-wave radar human activity recognition method based on a contrastive learning network, CLHAR-Net. This method utilized unlabeled samples to extract human activity micro-Doppler features through self-

improvised pre-training and fine-tunes the classifier with a small number of labeled samples to achieve high-precision millimeter-wave radar human activity recognition. Changlong Wang et al. [7] proposed a method, MIMO-3DPose, implementing three-dimensional multi-person pose estimation through walls using MIMO millimeter-wave radar. This method, through a cross-modal supervision learning pipeline, combined synchronized cameras and radar data and used an end-to-end Transformer network, ERPENT, to directly predict 3D poses from radar heatmaps.

This paper aims to review the current application status of wireless sensing technology in human pose estimation, analyze its technical challenges, and anticipate future development directions. This paper will introduce the application of Wi-Fi signals and millimeter-wave radar in human pose estimation, conduct a comparison and analysis, dissect the challenges of their future development, and give a perspective on future development.

COMPARATIVE ANALYSIS OF TYPICAL TECHNOLOGIES

TABLE 1: Wireless Sensing Human Pose Estimation Technology

Device	Feature Information	Representative Work
Wi-Fi	Channel State Information (CSI)	Widar 3.0
	Channel State Information (CSI)	GESFI
	Channel State Information (CSI)	Person-in-WiFi 3D
Millimeter-Wave Radar	Radar Point Cloud	Dynamic GNN-MB
	Micro-Doppler Map (MDM)	CLHAR-Net
	Dual-Channel Horizontal/Vertical Radar Heatmap I_h 、 I_v	MIMO - 3DPose

Table 1 presents the two typical wireless sensing technologies and their representative works analyzed in this paper, as well as the corresponding feature information used.

Human Pose Estimation Based on Wi-Fi

The channel state information (CSI) of Wi-Fi signals contains the changes of amplitude and phase of the signals during transmission, which can reflect human body movements in the environment. By analyzing the CSI data, human pose estimation can be achieved.

In this section, this paper will introduce three representative works on human pose estimation based on Wi-Fi signals: the Widar 3.0 system [1, 2], the GESFI system [3], and the Person-in-WiFi system [4]. These three systems use the channel state information (CSI) of Wi-Fi signals as feature information for human pose estimation.

Widar 3.0 System

Yi Zhang et al. [1] proposed a novel human action recognition system, Widar 3.0. By extracting domain-independent characteristic body-coordinate velocity profile (BVP), this system achieves gesture recognition adaptable to different environments with only one-time training. This system utilizes the CSI signals from commercial Wi-Fi devices and recovers BVP from multi-link signals through compressive sensing technology. It also designs a deep learning model that fuses spatiotemporal features. Experiments show that Widar 3.0 achieves accuracies of 92.4%, 89.7%, 82.6%, and 88.9% in cross-environment, cross-position, cross-direction, and cross-user scenarios, respectively, significantly outperforming the original CSI (40.2%) and DFS (77.8%) methods, and it can adapt to new domains without re-training.

However, the Intel 5300 wireless NIC and multi-transceiver hardware facilities adopted in the construction of this system are relatively expensive. Moreover, this research focuses on the study of algorithms and technologies and is only at the research stage.

For this, Liang Fang et al. [2] used an ESP32 microcontroller and an external antenna to collect Wi-Fi signals as the data acquisition module to collect CSI data, significantly reducing the hardware cost, and they combined the deep learning model cross-domain gesture recognition system Widar 3.0 for gesture recognition. Subsequently, this action

recognition system was deployed on the host computer and placed in the home area to achieve real-time control of the smart home, solving the interaction problem of this system in practical applications and enhancing the user experience.

The design of this system utilized the Widar 3.0 dataset, which includes three indoor scenarios and 22 dynamic gestures. This system can offer users better privacy protection and personalized interaction methods, but it has issues such as insufficient recognition accuracy, stability, and real-time performance, as well as a lack of algorithms for recognizing actions in multi-person scenarios.

GESFI System

Xiang Zhang et al. [3] proposed an objective gesture recognition system based on Wi-Fi, named GESFI, aiming to address the issue of limited cross-domain generalization ability caused by the reliance on subjective domain labels in traditional methods. For the collected CSI data, the system denoises it with the CSI-ratio method, extracts Doppler Frequency Shift (DFS) and phase information through short-time Fourier transform, and visualizes the data. A convolutional neural network (CNN) conducts two rounds of rough training for pre-learning in the gesture classification task, enabling the model to understand the data distribution initially. Then, a self-supervised pseudo-labeling strategy is adopted to mine latent domain labels, and adversarial training is utilized to minimize the differences between domains and learn domain-invariant features.

Experiments show that GESFI performs well on multiple datasets. On the Widar3.0 dataset (6 pairs of antennas), the accuracy rates for cross-location, cross-environment, and cross-user tasks reached 99.26%, 99.32%, and 99.37% respectively; on the ARIL dataset (single pair of antennas), the accuracy rate for the cross-location task was 75.52%; on the XRF55 dataset, the accuracy rates for the cross-environment and cross-user tasks were 62.15% and 67.18% respectively. In real-world environment tests, the accuracy rates for cross-position and cross-direction were 46.00% and 42.89%, respectively. These data indicate that GESFI can effectively improve the accuracy and generalization ability of gesture recognition in cross-domain scenarios.

This system enhances the generalization ability of cross-domain gesture recognition through objective domain division, but it also faces shortcomings, such as the difficulty in determining the number of latent domains and the limited handling of continuous changes in domains.

Person-in-WiFi 3D System

Current Wi-Fi methods have made progress in single-person 2D/3D and multi-person 2D pose estimation, but multi-person 3D pose estimation remains an unsolved challenge. Traditional methods encounter problems such as large network scale, slow training, and inefficient post-processing when extended to 3D.

In response to this, Kangwei Yan et al. [4] proposed a Wi-Fi-based multi-person 3D human pose recognition system called Person-in-WiFi 3D. This system achieves breakthroughs mainly through the following two aspects: first, it deploys three Wi-Fi receivers to enhance the ability to capture spatial reflection signals; second, it implements end-to-end pose estimation based on the Transformer framework, including a CSI encoder, a pose decoder, and a refined decoder. The system tokenizes the CSI signals and extracts spatiotemporal features, which are processed through a multi-head self-attention module and feed-forward network, and ultimately outputs 3D poses coordinates. Training is conducted using a set-based Hungarian loss combined with focal loss and mean squared error optimization.

The experiment was conducted on a self-built dataset. This dataset used the collected CSI data and RGB-D videos captured by the Azure Kinect camera and synchronized as the data, containing over 97,000 samples from 7 volunteers in different scenarios. The results showed that the 3D joint positioning errors of the system in single-person, two-person, and three-person scenarios were 91.7mm, 108.1mm, and 125.3mm, respectively, with an MPJPE (Mean Per Joint Position Error) of 107.2mm, comparable to that of cameras and millimeter-wave radars. Ablation experiments verified the crucial roles of the number of receivers, refined decoder, and phase denoising. This system provides a new scheme for human pose estimation in privacy-sensitive scenarios.

This system performs exceptionally well in single, double, and triple-person scenarios and demonstrates highly efficient storage and computing capabilities. However, it is also constrained by the performance of the labeling tools and issues related to space configuration.

Human Pose Estimation Based on Millimeter-Wave Radar

Millimeter-wave radar generates point cloud data or radar heatmaps by emitting high-frequency electromagnetic waves and receiving reflected signals and then estimates human pose.

In this section, three representative works on human pose estimation based on millimeter-wave radar will be introduced: Dynamic GNN-MB [5], CLHAR-Net [6], and MIMO-3DPose [7]. These three methods respectively utilize radar point cloud data, micro-Doppler maps (MDM), and dual-channel horizontal/vertical radar maps I_h, I_v as feature information for human pose estimation.

Dynamic GNN-MB Network

Guoliang Peng et al. [5] proposed a human action recognition method of millimeter-wave radar based on Dynamic GNN-MB network.

For the sparse point cloud generated by millimeter-wave radar, this system employs a graph neural network (Dynamic GNN) with a dynamic edge selection function, which automatically learns edge weights and extracts features based on the features of adjacent vertices without the need for manual edge specification. Then, it uses an MLP network to adjust the dimensions (where the Max function selects the most representative information), combines stacked bidirectional gated Recurrent Units (BIGRU) to capture the temporal dynamic features and context relationships in the point cloud data, and finally outputs the recognition results through a fully connected layer.

The experiment was validated using the public dataset MMActivity. The results indicate that the accuracy of this model in human action recognition reaches 97.05%, which is 3.43% higher than that of Point-GNN + Bidirectional GRU and represents a significant improvement compared to baseline methods (such as 90.47% of CNN + Bi-LSTM). The study points out that this approach effectively processes sparse data by dynamically adjusting the graph structure, and the bidirectional temporal modeling capability of BIGRU enhances feature representation, offering a new solution for high-precision action recognition under privacy protection.

This system is capable of effectively handling the sparse point cloud data produced by millimeter-wave radar, and it possesses advantages such as high recognition accuracy and excellent privacy protection. Nevertheless, its performance in real and complex interference environments remains doubtful, and it has high demands for computing resources.

CLHAR-Net System

Haoze Du et al. [6] proposed a millimeter-wave radar human activity recognition method based on a contrastive learning network, CLHAR-Net. This system combines self-supervised pre-training with supervised fine-tuning to address the issue that traditional methods rely on many labeled samples.

The system initially conducts processing on the original radar echo data through differentiation, Fast Fourier Transform (FFT), and Short-Time Fourier Transform (STFT), and eventually concatenates along the time dimension to generate the Micro-Doppler Map (MDM) as the input. In the self-supervised pre-training phase, unlabeled samples are used, and a dictionary query task is constructed via a dual-encoder structure (the basic encoder and the momentum encoder). The InfoNCE loss function is employed to compel the features of positive samples to approach each other and those of negative samples to distance apart, thereby learning the deep representation of the MDM. After pre-training, the parameters of the convolutional encoder are frozen, and only a small quantity of labeled samples are utilized to fine-tune the downstream classifier (comprising two fully connected layers and a Softmax layer), ultimately attaining high-precision classification.

The experiment is based on the data of five types of activities (walking, running, boxing, squatting, and swinging) collected by the Ti-2243 radar (550 samples for each type). The results show that the CLHAR-Net achieves an accuracy rate of 99.63% with 100% labeled samples and significantly outperforms the traditional CNN in the low-label scenarios of 5% and 10%, verifying its efficiency and generalization ability. This system utilizes unlabeled samples and has the advantages of excellent classification performance and strong generalization ability. However, it consumes a large amount of computing resources in the pre-training stage and relies on data augmentation methods.

MIMO-3DPose

Changlong Wang et al. [7] proposed a method called MIMO-3DPose for implementing real-time through-wall multi-person 3D pose estimation based on MIMO radar. This system employs a cross-modal supervised learning framework to achieve human pose estimation in obstructed scenarios by synchronizing a camera with a self-developed MIMO radar.

The core of the system consists of three parts: Firstly, 1-2GHz MIMO radar (a 4-transmitter and 16-receiver antenna array, center frequency 1.5GHz, bandwidth 1GHz) and a 30fps camera are used to synchronously collect data. The 3D pose extracted from the RGB image is transformed into the world coordinate system with the radar as the origin through coordinate transformation. Secondly, Moon et al.'s monocular 3D pose estimation method is adopted as the teacher network. YOLO is used to detect the human bounding boxes, RootNet to locate the human body center, and PoseNet to output the absolute 3D human pose, generating supervisory signals. Finally, an end-to-end student network ERPENT based on Transformer is designed. It takes dual-channel (horizontal/vertical) radar heatmaps (integrating static limb reflection and dynamic speed information) as input and directly predicts the 3D pose candidate set (14 key point coordinates) and person classification labels through a multi-feature fusion module and a feature encoding-decoding module.

The dataset employed in the experiment was self-acquired, totaling over 1,160,000 pairs of data from more than 12 participants. Among them, 1,096,000 frames were utilized for training and 64,000 frames for validation. The experiment indicates that in the 28-cm brick wall scenario, the average positioning error of ERPENT-D is 40.15 mm, with a single-frame processing time of 0.082 seconds, significantly superior to RF-Pose3D (43.67 mm) and the method proposed by Song et al. (48.36 mm).

This system features strong real-time performance, high precision, strong wall-penetrating capability, and reconfigurable for occluded postures; however, the experimental scenarios are limited, and there are issues with insufficient precision for specific individuals (such as those with tall height or heavy weight).

Technical Comparison

TABLE 2: Comparison of Wi-Fi Signals and Millimeter-Wave Radar Performance

Device	Accuracy	Precision	Recall	F1 Score
Wi-Fi Signals	65.09%	67.91%	65.09%	65.72%
Millimeter-Wave Radar	97.78%	97.99%	97.68%	97.78%

Ajaya Dahal et al. [8] compared the performance discrepancies of millimeter-wave radar and Wi-Fi-CSI technology in seven types of human activity recognition (HAR). The experiment employed 77 GHz FMCW radar (INRAS Radarbook2) and a Raspberry Pi 3B+ with Nexmon firmware to collect data of seven activities, such as falls and running in a unified scene, and constructed a public dataset containing 700 samples. Micro-Doppler signatures were generated through time-frequency analysis, and a 2D CNN model was adopted for classification. The experimental outcomes encompassed four indicators: accuracy, precision, recall, and F1 score (as shown in Table 2).

Evidently, concerning the parameters reflecting recognition performance, such as accuracy, precision, recall, and F1 score, millimeter-wave radar is conspicuously superior to Wi-Fi signals. Nevertheless, the deployment cost of millimeter-wave radar is considerably higher than that of Wi-Fi [9,10]. Hence, in low-cost deployment contexts such as smart homes, it is more appropriate to select Wi-Fi signals for human pose estimation; in scenarios with high-precision requirements such as factories and medical fields, choosing millimeter-wave radar is more suitable.

APPLICATION ANALYSIS

At present, Wi-Fi signals, with their extensive deployment and non-intrusive characteristics, are mainly utilized in the following scenarios in human pose estimation:

Firstly, in smart home and interactive control. Utilizing the channel state information (CSI) of Wi-Fi, the system can recognize user gestures by analyzing the signal reflection characteristics, enabling non-contact control of home appliances, lights, and other devices. For example, the Widar3.0 system has achieved a gesture recognition accuracy rate of 92.4% in cross-environment and cross-user scenarios by extracting the body-coordinate velocity profile (BVP) of human movements and is applicable to scenarios such as smart home and virtual reality (VR) that require natural interaction[11-13].

Secondly, in health monitoring and elderly care. Wi-Fi sensing can be employed for real-time monitoring of the daily activities of the elderly, such as fall detection and sleep quality analysis. For example, the Person-in-WiFi 3D system, via multi-antenna deployment and Transformer networks, has accomplished multi-person 3D pose estimation [14]. In the elderly care environment, it can provide early warnings of health risks by monitoring posture changes while safeguarding user privacy.

Thirdly, in cross-domain adaptability scenarios. The superiority of Wi-Fi technology resides in its capacity to adapt to new environments without the necessity of re-training. For example, the GESFI framework has maintained a recognition accuracy of above 95% in cross-location and cross-direction scenarios by objectively mining the latent domain features in signals [3], rendering it applicable to diverse indoor environments such as offices and classrooms.

Millimeter-wave radar, with its high resolution and penetration capability, has an irreplaceable advantage in human pose estimation in complex environments:

Firstly, in security and surveillance. Millimeter-wave radar is capable of penetrating obstacles such as walls and smoke, facilitating covert monitoring. For example, the ERPENT system, by means of MIMO radar and Transformer networks, has achieved a joint positioning accuracy of 40.15mm in a 28cm brick wall scenario [7] and is applicable to scenarios such as counterterrorism and post-disaster rescue.

Secondly, in industrial and medical scenarios. In the industrial environment, millimeter-wave radar can be utilized for monitoring the postures of workers to prevent accidents; in the medical domain, it can be employed for analyzing the movements of rehabilitation training. For example, a system based on the Dynamic Graph Neural Network (Dynamic GNN-MB) attains an accuracy rate of 97.05% in human action recognition by processing sparse point cloud data, being applicable to precise medical scenarios [15].

Thirdly, multi-target tracking in complex circumstances. The human pose estimation technology based on MIMO millimeter-wave radar sparse point cloud data holds broad application prospects in the domain of multi-target tracking in complex environments[16,17]. Through the combination of the clustering method for human tracking, this technique can effectively cope with interferences such as occlusion and illumination variations, achieving high-precision recognition for single and double persons, with the error probabilities reduced by 23.4% and 31.1%, respectively [18]. It has extensive application scenarios in complex situations with multiple targets, such as smart homes, elderly care and medical services, and industrial production.

In the future, human pose estimation based on wireless sensing will have a wide range of application scenarios.

In the field of smart homes, Wi-Fi technology achieves contactless gesture control of home appliances through CSI signal analysis and combines cross-domain adaptability to solve the problem of multiple-person motion conflicts. In medical and health care, the high precision of millimeter-wave radar is suitable for fall detection, and combined with physiological signals, it can enhance the reliability of early warning; Wi-Fi, with its low cost, supports the monitoring of daily activities. In virtual reality, the real-time performance and sub-centimeter accuracy of millimeter-wave radar meet the interaction requirements, and the low latency optimization or technological updates of Wi-Fi can build a multi-modal, low-cost, and low-latency solution. In emergency rescues, the wall-penetrating capability of millimeter-wave radar enables the rapid location of survivors in ruins, and the wide-area coverage of Wi-Fi can achieve multi-region coordinated search and rescue.

ANALYSIS OF CHALLENGES AND FUTURE PROPECTS

In the research of applying wireless sensing technology to human pose estimation, numerous multi-dimensional technical bottlenecks await urgent breakthroughs: Firstly, there are marked limitations in accuracy and multi-user scenarios. For Wi-Fi signals, the positioning error increases conspicuously in multi-user scenarios compared to single-user ones, while millimeter-wave radar contends with the superimposition of complex human motion signals. Secondly, real-time performance and latency constrict the dynamic interaction experience. Wi-Fi systems suffer from action recognition delays, and millimeter-wave radar needs to optimize the single-frame processing time to the millisecond level. Thirdly, the signal resolution bottleneck is pronounced. The inadequate spatial resolution of Wi-Fi leads to the misidentification of similar actions, and the low resolution of radar impacts the accuracy and dynamic stability of target recognition. Fourthly, complex environmental interferences are notable. Issues such as multipath effects, human occlusion, electromagnetic noise, and the attenuation of millimeter waves when penetrating walls reduce the robustness of the system. Fifthly, data dependency and generalization predicaments are prominent. The cost of synchronizing camera annotations is high, and there are privacy risks. The subjective domain labels of Wi-Fi restrict the generalization ability across scenarios, and radar needs to address the issue of cross-modal spatiotemporal alignment. Finally, the constraints of computing resources are significant. The Transformer model of Wi-Fi and complex models like ERPENT of radar pose challenges to the deployment on edge devices, demanding further lightweight design and edge computing optimization.

In the future, the development of wireless sensing technology in the field of human pose estimation can be deeply advanced around the following three core directions:

Under the empowerment of 6G technology, wireless sensing technology has broad prospects in the field of human pose estimation. The synergy of reconfigurable intelligent surfaces [8] and the terahertz frequency band [9], where the terahertz waves possess high bandwidth and short wavelength [9], combined with the beam control capability of reconfigurable intelligent surfaces [8], can enhance signal coverage and positioning accuracy in complex environments, enabling centimeter-level recognition and addressing the problem of non-line-of-sight occlusion. The collaboration of edge AI [10] and computing power networks [11] brings a low-latency and high-concurrency real-time processing experience. When combined with federated learning [16], it improves inference efficiency and protects privacy, being applicable to scenarios such as real-time monitoring. Semantic communication networks [12] revolutionize the data transmission mode, focusing on the transmission of meaningful information, reducing redundancy, and improving spectral efficiency. Simultaneously, in conjunction with 6G dynamic spectrum management [19], it optimizes resource allocation and supports real-time interaction among multiple devices.

Algorithm optimization, as a major direction for future advancement, can center on multimodal fusion, lightweight model design, and self-supervised learning. Regarding multimodal fusion, the synergy between millimeter-wave radar and visual sensors (such as cameras and LiDAR) [13] can be accomplished through data-level, feature-level, and decision-level fusion strategies. This integration exploits the penetrability and all-weather stability of radar along with the high-resolution spatial information of visual sensors, enhancing the accuracy of pose estimation in complex scenarios and compensating for the limitations of individual sensors. Lightweight model design merits attention. Existing studies, such as LMFormer [14], introduce a lightweight multi-feature perspective Token Mixer and a multi-scale information propagation mechanism, significantly reducing model complexity while maintaining high accuracy. Self-supervised learning [15] utilizes large-scale unlabeled wireless signal data for pre-training, facilitating the model to acquire robust features and enhance the performance of small-sample supervised learning, thereby resolving the issue of data annotation.

Privacy protection is of great significance. Federated Learning (FL) [16], as a key technology, can concentrate on the following aspects in the future: integrating Differential Privacy (DP) and Homomorphic Encryption (HE) to strengthen data protection; optimizing communication protocols to lower the energy consumption and bandwidth demands of edge devices and address communication bottlenecks; exploring algorithms adapted to the independent and identically distributed (non-IID) data to enhance the generalization ability of the model; developing lightweight frameworks to be compatible with low-computing-power devices and facilitate the deployment of the Internet of Things (IoT); establishing standardized benchmark datasets and evaluation systems to promote cross-domain collaborative development and offer superior privacy protection solutions for wireless human pose estimation.

The above are three aspects that merit attention in the development process of wireless human pose estimation technology. Progress in these three aspects is anticipated to improve the performance of human pose estimation technology and user experience and contribute to social advancement.

COCLUSIONS

This paper conducts a systematic literature review and technical comparison analysis to deeply explore the current application status, typical technologies, and development potential of Wi-Fi signals and millimeter-wave radar in human pose estimation.

This review focuses on Wi-Fi signal-based wireless sensing technology and millimeter-wave radar-based wireless sensing technology, each with three representative works introduced and analyzed. For Wi-Fi signal-based human pose estimation technology, this paper analyzes Widar3.0 (cross-domain gesture recognition through BVP features), GESFI (mining potential domain features based on self-supervised pseudo-label strategy), and Person-in-WiFi 3D (multi-person 3D pose estimation using Transformer), all of which are based on channel state information (CSI), revealing their advantages in low-cost deployment and privacy protection, as well as the limitations of multipath effects and insufficient real-time performance. For millimeter-wave radar-based human pose estimation technology, this paper discusses Dynamic GNN-MB (processing sparse point clouds with dynamic graph neural networks), CLHAR-Net (contrastive learning combined with unlabeled data), and MIMO-3DPose (through-the-wall localization with cross-modal supervision learning) around point clouds, micro-Doppler maps (MDM), and radar heat maps respectively, highlighting their high precision and penetration advantages, as well as the challenges of high cost and robustness in complex scenarios.

The two technologies exhibit complementarity in application scenarios: Wi-Fi applies to privacy-sensitive smart home and health monitoring, while millimeter-wave radar is better-suited for scenarios demanding high precision such

as security and industry. Currently, technical bottlenecks are mainly focused on multi-user interference, real-time optimization, robustness in complex environments, and data dependence, among others.

Future research could be centered on the following orientations: Firstly, based on the terahertz frequency band and reconfigurable intelligent surfaces of 6G technology, overcome the signal occlusion issue in non-line-of-sight scenarios and attain centimeter-level positioning precision. Secondly, enhance the algorithmic efficiency and generalization capability through multi-modal fusion (such as radar and visual sensors) and lightweight model design (such as LMFormer). Lastly, utilize federated learning techniques to boost the performance of privacy protection, which can be realized by means of combining differential privacy and homomorphic encryption, optimizing communication protocols, exploring federated learning algorithms adapted to independent and identically distributed data, developing lightweight federated learning frameworks, and establishing standardized federated learning benchmark datasets and evaluation systems. These advancements will offer more dependable technical support for domains such as intelligent interaction, telemedicine, and emergency rescue, and contribute to the construction of a secure and efficient intelligent society.

REFERENCES

1. Zhang, Yi, et al. "Widar3. 0: Zero-effort cross-domain gesture recognition with Wi-Fi." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.11 (2021): 8671-8688.
2. Liang Fang, Quanwei Jiang, Sichao Li et al. "Research on smart home control system based on motion recognition"[J]. *Modern Computer*, 2024, 30(20): 114-116+120.
3. Xiang Zhang, Yuanhao Feng, Jingyang Huang, et al. Objective Gesture Recognition Based on WiFi. *TechRxiv*. December 02, 2024.
4. Yan, Kangwei, et al. "Person-in-wifi 3d: End-to-end multi-person 3d pose estimation with wi-fi." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
5. Guoliang Peng, Haoran Li, Fen Hu, et al. "Human Action Recognition Method of Millimeter-wave Radar Based on Dynamic GNN-MB Network"[J/OL]. *Modern Radar*, 1-10[2025-0308]. <http://kns.cnki.net/kcms/detail/32.1353.tn.20250115.1648.002.html>.
6. Du, Haoze, et al. "Millimeter wave radar human activity recognition with a contrastive learning network." *2023 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*. IEEE, 2023.
7. Wang, Changlong, et al. "Real-time through-wall multiperson 3D pose estimation based on MIMO radar." *IEEE Transactions on Instrumentation and Measurement* (2024).
8. Umer, Anum, et al. "Reconfigurable Intelligent Surfaces in 6G Radio Localization: A Survey of Recent Developments, Opportunities, and Challenges." *IEEE Communications Surveys & Tutorials* (2025).
9. Siddiky, Md Nurul Absar, et al. "A Comprehensive Exploration of 6G Wireless Communication Technologies." *Computers* (2025).
10. Gill, Sukhpal Singh, et al. "Edge AI: A taxonomy, systematic review and future directions." *Cluster Computing* 28.1 (2025): 1-53.
11. Yukun, Sun, et al. "Computing power network: A survey." *China Communications* (2024).
12. Guo, Shaolong, et al. "A survey on semantic communication networks: Architecture, security, and privacy." *IEEE Communications Surveys & Tutorials* (2024).
13. Wang, Shuai, et al. "Multi-modal fusion sensing: A comprehensive review of millimeter-wave radar and its integration with other modalities." *IEEE Communications Surveys & Tutorials* (2024).
14. Li, Biao, Shoufeng Tang, and Wenyi Li. "LMFormer: Lightweight and multi-feature perspective via transformer for human pose estimation." *Neurocomputing* 594 (2024): 127884.
15. Dash, Ankan, et al. "Self-Supervised Learning for User Localization." *2024 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2024.
16. Dritsas, Elias, and Maria Trigka. "Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications." *Journal of Sensor and Actuator Networks* 14.1 (2025): 9.
17. Dahal, Ajaya, et al. "Comparison Between Wi-Fi-CSI and Radar-Based HAR." *2024 IEEE Radar Conference (RadarConf24)*. IEEE, 2024.
18. Zhou, Xiaohua, et al. "Human body recognition based on the sparse point cloud data from MIMO millimeter-wave radar for smart home." *Multimedia Tools and Applications* 83.8 (2024): 22055-22074.
19. Lindenschmitt, Daniel, et al. "Dynamic Spectrum Management for 6G Network-in-Network Concepts." *2024 15th International Conference on Network of the Future (NoF)*. IEEE, 2024.