Research on CSI-based Gesture Recognition Technology and Their Typical Applications

Litao Wei

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100000, China
Wlt777@bupt.edu.cn

Abstract. In recent years, with the development of technology, gesture recognition technology based on Wi-Fi Channel State Information (CSI) has received widespread attention. This technology can realize non-contact, wireless perception recognition of gesture movements, and has many advantages such as high precision, good compatibility and strong environmental adaptability. This study firstly introduces the development background and research significance of CSI gesture recognition, highlighting its potential to revolutionize human-computer interaction by enabling seamless control of smart devices without the need for physical contact or dedicated sensors. Furthermore, it emphasizes the societal value of this technology in assisting individuals with mobility impairments, offering an accessible and intuitive interface for digital interaction. Then, it compares and analyzes eight typical gesture recognition systems including PAC-CSI, WiCGesture, WiGNN, WiGesFree, UbiGest, Wi-SFDAGR, WiDual, and CrossFi. It discusses their key technologies, performance and application potential, and analyzes the challenges they currently face, and finally looks at the future trends of these technologies.

Keywords: Wireless Sensing, Wi-Fi Channel State Information, Gesture Recognition, Human–Computer Interaction.

1 Introduction

With the development of technology, human-computer interaction has become an important bridge connecting users and intelligent systems. It is undergoing a transition from traditional input methods such as keyboard and mouse and touch screen to natural interaction methods. Gesture recognition is gradually becoming one of the key technologies in intelligent interaction systems. Traditional gesture recognition mostly relies on visual sensors, wearables, infrared or sound detection. However, these methods generally have significant limitations due to the influence of light, susceptibility to object occlusion, high cost and small detection range. In recent years, CSI in Wi-Fi signals has become an emerging means of contactless sensing. CSI has become an important breakthrough in gesture recognition due to its high accuracy, ability to sense through walls, strong environmental adaptability and low cost. Therefore, CSI-based gesture recognition technology has rapidly become a research hotspot, and is widely used in

S. Zhang (ed.), *Proceedings of the 2025 International Conference on Electronics, Electrical and Grid Technology (ICEEGT 2025)*, Advances in Engineering Research 292,
https://doi.org/10.2991/978-94-6463-986-5_30

smart home, virtual reality (VR) and augmented reality (AR), in-vehicle control, gaming, and other scenarios. In recent years, multiple studies have further advanced the field. For example, the PAC-CSI system proposed by Jian Su et al. realizes high-precision gesture recognition in a digital twin environment through a phase-based attention mechanism [1]. In addition, Ruiyang Gao et al.'s WiCGesture system successfully solves the problem of recognizing continuous gestures by introducing “metamotion” features, and its recognition accuracy reaches 89.6% in complex gesture scenarios [2]. WiGNN system proposed by Yinan Chen and XiaoXia Huang enhances the accuracy of cross-domain gesture recognition by graph neural network (GNN) and achieves high recognition rate in multiple environments [3]. This study systematically reviews the development status of CSI-based gesture recognition technologies and their typical technologies and core algorithms. Then, it summarizes application scenarios and discusses key challenges and future trends.

2 Typical technology comparison and analysis

2.1 PAC-CSI

PAC-CSI is a novel gesture recognition system based on CSI designed for digital twin environments. The system addresses performance degradation across domains by processing CSI data, eliminating noise, and applying a lightweight deep neural network with a phase-based attention mechanism. With 99.46% recognition accuracy within a single domain, PAC-CSI also performs well across locations, directions, users and environments. On the plus side, PAC-CSI is able to maintain a high level of recognition accuracy across environments and users. PAC-CSI's cross-domain adaptability enables it to respond effectively to changes in different locations, users and environments. The lightweight network architecture of the PAC-CSI allows for fast processing of gesture data in a real-time environment. On the downside, PAC-CSI doesn't support user identification, so it can't identify specific users. The sensitivity of the system to ambient noise may affect accuracy in the presence of poor signal quality. Additionally, system performance depends on the quality of the training data. Therefore, more data augmentation may be necessary to improve results when insufficient data is available. The application areas of PAC-CSI mainly include digital twin environments such as smart home control, in-vehicle interaction, VR and AR [1].

2.2 WiCGesture

WiCGesture, a continuous gesture recognition system based on Wi-Fi signals, was studied by Ruiyang Gao et al. It aims to solve the problem of the need to pause when processing continuous gestures. The key to this technique is the introduction of “metamotion” features to characterize the geometry of gestures, which decompose a continuous flow of gestures into elementary motion segments. WiCGesture uses changes in the dynamic phase of the Wi-Fi signal to capture the characteristics of hand movements and can effectively recognize complex continuous gestures. On the plus side, WiCGesture has high recognition accuracy, reaching 89.6% (for the set of numbers) and 88.3% (for the set of Greek letters) in successive gesture scenarios. For different users and gesture speeds, WiCGesture is more adaptable and can maintain efficient recognition in multiple environments with good stability and robustness. On the downside, WiCGesture may misrecognize certain gestures, especially if they are too simple or too similar. In addition, its recognition accuracy decreases for extremely fast hand movements. WiCGesture's applications include smart home control, virtual reality, augmented reality, and HCI scenarios. It is especially suited for environments requiring natural, touchless interaction, such as continuous text or numeric input[2].

2.3 WiGNN

WiGNN is a cross-domain gesture recognition system based on Wi-Fi signals proposed by Yinan Chen and XiaoXia Huang et al. The system improves the accuracy of cross-domain gesture recognition by processing CSI, which is collected by multiple receivers, via GNN. The GNN combines causal and multi-scale convolutions to extract temporal and frequency features, respectively. WiGNN achieves an average accuracy of 94.0% in cross-domain tasks. On the plus side, WiGNN effectively captures the spatio-temporal dependencies between different receivers and improves the accuracy of cross-domain gesture recognition. WiGNN can extract time-frequency features from CSI signals. This solves the problem of vanishing gradients that traditional methods may encounter when working with time-series data. Additionally, the Dynamic Topology Aggregation Layer (DTAL) enables WiGNN to efficiently aggregate features from different receivers, ensuring the model's strong cross-domain adaptability. WiGNN also improves robustness in different user situations by using direction-independent data enhancement techniques. On the downside, WiGNN is dependent on multiple receivers when the number of receivers is unstable or unevenly distributed. Application areas for WiGNN include smart home, health monitoring and human-computer interaction. WiGNN is particularly suitable for gesture recognition tasks in low-cost and contact-free environments [3].

2.4 WiGesFree

WiGesFree is a gesture recognition system based on Wi-Fi signals proposed by Ding et al. It aims to improve accuracy, reduce dependence on data samples and devices, and eliminate the need for model training. The technique's key is combining the sample point distance (SPD) algorithm and dynamic phase change to extract gesture features and utilizing the CSI ratio for feature extraction and endpoint detection, which reduces noise interference. WiGesFree improves gesture recognition accuracy across locations, scenarios and users, averaging 93.67%. On the plus side, WiGesFree is able to achieve high gesture recognition accuracy in the absence of a large number of data samples and sensor devices. WiGesFree accurately extracts the start and end points of gestures, improving recognition accuracy. Additionally, WiGesFree mitigates noise interference effectively and improves recognition performance. On the downside, WiGesFree may suffer from signal fading and multipath effects in complex environments, resulting in reduced recognition accuracy in certain scenarios. Additionally, brief pauses in gestures or environmental disturbances may affect WiGesFree's accuracy in recognizing gestures, especially fast ones. WiGesFree's applications include smart homes, health monitoring, touch interfaces, virtual reality, and entertainment [4].

2.5 UbiGest

UbiGest is a smartphone-based gesture recognition system utilizing CSI proposed by Seung-Hyun Jeong et al. The key technology solves the problem of performance degradation in traditional Wi-Fi gesture recognition systems caused by environmental changes and changes in user position and orientation by using only beacon frames sent from surrounding access points (APs). On the plus side, UbiGest uses CSI from smartphones and multiple APs for gesture recognition, which effectively extends the recognition range and avoids the limitations of traditional methods that require multiple transmitters. UbiGest maintains high accuracy in new environments, especially with a limited number of APs. Additionally, the system has a high degree of environmental adaptability. By selecting appropriate channels and APs, it can effectively mitigate the effects of user position, attitude and environmental changes. On the downside, the performance of UbiGest may suffer in environments with a low number of APs or strong interference, resulting in decreased recognition accuracy. Additionally, the accuracy of UbiGest's recognition decreases when strong interference is present. UbiGest's applications include smart homes, health monitoring, location-aware systems, and HCI. It is especially suitable for achieving full-coverage gesture recognition without wearing a device [5].

2.6 Wi-SFDAGR

Wi-SFDAGR is a Wi-Fi cross-domain gesture recognition system proposed by Huan Yan et al. It aims to address the limitation of traditional methods that rely on source domain data. The system improves target domain feature extraction using attraction-diffusion networks and local neighborhood uncertainty estimation. This allows the system to support cross-domain adaptation when source data is unavailable. In terms of advantages, Wi-SFDAGR proposes a domain adaptation method in case of source data unavailability, which effectively solves the data privacy problem. Wi-SFDAGR can improve the model's cross-domain generalization ability and recognition accuracy. Wi-SFDAGR achieves 97.30%, 97.17%, and 95.52% accuracy in cross-position, cross-direction, and cross-environment tasks, respectively. Additionally, Wi-SFDAGR can rely only on pre-trained models for target-domain adaptation without source-domain data, which reduces computational and storage overheads. On the downside, the recognition accuracy of Wi-SFDAGR decreases when the environment varies greatly. Additionally, the high computational complexity of attraction-diffusion networks and local neighborhood uncertainty estimation may impact real-time processing capabilities. Wi-SFDAGR has many application areas, including smart homes, health monitoring, and

VR. It is particularly well-suited for contactless gesture recognition tasks that require handling across environments, locations, and directions [6].

2.7 WiDual

WiDual is a real-time dual-task recognition system based on Wi-Fi signals proposed by Chenhong Cao et al. It aims to simultaneously realize gesture recognition and user identification across domains. The key to the system's technology is the use of CSI visualization to convert Wi-Fi signals into images, thus simplifying the feature extraction process and improving the efficiency of model training. WiDual combines spatial and channel attention mechanisms to adaptively extract key features of gestures and user identity. WiDual also fuses the features of the two through a collaboration module to improve the accuracy of dual-task recognition. On the plus side, WiDual is able to efficiently extract cross-domain features, which effectively improves cross-domain accuracy. WiDual is capable of simultaneous gesture recognition and user identification, increasing the system's versatility. Additionally, the Collaboration Module from WiDual enhances dual-task recognition performance and reduces interference between tasks. On the downside, WiDual's recognition accuracy may be reduced with a smaller number of receivers or large changes in location. Additionally, the computational complexity of WiDual's attention mechanism and collaboration module may be challenging for some resource-constrained devices. WiDual's application areas include smart home, health monitoring, and VR [7].

2.8 CrossFi

CrossFi is a Wi-Fi cross-domain awareness framework based on Siamese network proposed by Zijian Zhao et al. to solve the Wi-Fi signaling domain transfer problem. Its core technologies include CSi-Net and Weight-Net. CSi-Net improves the similarity computation of traditional Siamese networks through a multi-head attention mechanism and is able to improve the accuracy of gesture and person recognition in cross-domain tasks. Weight-Net, on the other hand, generates category templates to improve cross-domain adaptation. On the plus side, CrossFi enables efficient cross-domain recognition in few-sample (91.72%) and zero-sample (64.81%) scenarios, especially in the absence of source domain data. CrossFi can adaptively generate templates for each category, thereby improving the model's accuracy in recognizing objects in different environments. Additionally, CrossFi is highly adaptable to cross-domain tasks and can perform gesture and person recognition in various scenarios. On the downside, CrossFi is still subject to signal variations and domain differences in complex environments, especially when there are large differences in the distribution between the target and source domains. In addition, CrossFi's performance may suffer when data is scarce or of low quality in the target domain. CrossFi's application areas include smart home, health monitoring, VR, etc. It is especially suitable for gesture recognition scenarios that require low-cost and contactless gesture recognition [8].

Table 1. Advantages and Disadvantages of Eight Typical Technologies

Typical Technologies	**Advantages**	**Disadvantages**
PAC-CSI	high accuracy strong cross-domain adaptability low time complexity	no support for user identification sensitive to environmental noise limited quality of training data
WiCGesture	ability to recognize continuous gestures high accuracy adaptable to different users and environments	misrecognition of simple gestures and very fast gestures
WiGNN	high accuracy ability to extract time-frequency features strong cross-domain adaptability robustness improvement	dependency on multiple receivers
WiGesFree	low dependency on equipment and samples high accuracy and cross-domain adaptability effective noise mitigation	low accuracy in complex environments sensitive to short pauses
UbiGest	wide range of recognition high accuracy and high robustness strong environmental adaptability	dependence on the number of APs sensitive to environmental noise
Wi-SFDAGR	data privacy concerns addressed high accuracy low computation and storage overhead	sensitive to complex environments high computational complexity
WiDual	high cross-domain accuracy high versatility dual-task recognition of gestures and identity	sensitive to receiver counts and variations high processing complexity
CrossFi	high accuracy with few and zero samples high flexibility strong cross-domain adaptability	sensitive to complex environments reliance on high-quality data

As shown in Table 1, a comparison of eight typical technologies in terms of advantages and disadvantages clearly demonstrates the characteristics of each. Each technology has its own unique advantages, but there are also some limitations

3 Application

In the field of intelligent HCI, CSI-based gesture recognition can provide a contactless interaction for smart home systems. Users control home appliances, lights and temperature control devices with gestures to enhance the convenience of life and intelligent experience. For example, Tulli et al. explored the application of gesture recognition technology in the smart home and proposed a low-cost gesture control scheme. This not only simplifies the interaction process but also improves recognition accuracy through deep learning methods [9]. At the same time, with the continuous development of equipment, these systems are gradually realizing seamless connectivity with different home appliances, further enhancing the quality of life of users.

Additionally, studies have shown that gesture recognition based on Wi-Fi signals has a wide range of potential applications in VR and AR, especially for high-precision motion capture via CSI without the need to wear a device [9, 10]. Without the need to wear a device, Users can naturally interact with virtual environments, in a way that not only enhances user immersion, but also opens up possibilities for future seamless interaction technologies.

In the in-vehicle system, the driver can control the car's entertainment system, navigation system and air conditioning with gestures. This contactless interaction improves driving safety, especially during driving, by eliminating the need for the driver to frequently operate on-board controls. The technology also monitors driver behavior in real time. By analyzing the driver's gesture movements, the system can detect the danger of fatigued or distracted driving and alert the driver through an alarm system. For example, the WiCGesture system analyzes the driver's hand gestures and facial expressions to provide timely alerts to remind the driver to stay vigilant [2].

In the gaming space, players can control characters or objects in the game through gestures, providing a more immersive gaming experience. For example, using Wi-Fi signals to recognize changes in a player's gestures can enable controller-independent game control. In multiplayer games, Wi-Fi gesture recognition can track the gestures of multiple players simultaneously. This allows players to not only control their characters freely, but also collaborate or compete with other players, enhancing the social and fun aspects of the game [9, 10]. With this technology, the interaction between players becomes more natural and fluid, and the entertainment and convenience of the game is enhanced.

4 Challenges and future directions

The issue of cross-domain adaptation is the primary challenge at hand. Since Wi-Fi signals are affected by a variety of environmental factors such as user position, attitude changes, and obstacles during propagation, the performance of the system tends to degrade significantly in different scenarios or user states. Therefore, how to achieve stable and accurate gesture recognition in changing environments remains a great challenge. Additionally, the less-sample learning problem also constrains the performance of the system. Sample scarcity has become the norm due to the high cost of acquiring high-quality labeled data. In this case, how to train a model with good generalization ability through limited data to improve the recognition accuracy and avoid overfitting is a key point where the current technology needs further breakthrough. Finally, the real-time and resource consumption issues of the system can’t be ignored. When dealing with complex gesture categories or multiple target users, there is a need to ensure that the system remains real-time and avoids computational resource and network bandwidth constraints.

In the future development, it is firstly necessary to introduce adaptive learning and sample-less learning mechanism to maintain efficient learning ability and quickly adapt to different environments and devices. Secondly, privacy protection mechanisms will be at the core of securing user data to ensure that the gesture recognition system can provide accurate services without violating user privacy. Additionally, the combination of edge computing and low-latency optimization will further enhance the accuracy and real-time performance of CSI-based gesture recognition technology. Finally, broad device compatibility can be achieved by lowering system deployment costs and reducing the need for specialized hardware.

5 Conclusion

CSI-based gesture recognition technology has developed rapidly in recent years, and has gradually become an important research direction in the field of human-computer interaction by virtue of its non-contact, low-cost, and easy-to-deploy features. This study systematically reviews and compares eight typical CSI-based gesture recognition systems, demonstrating the innovative path and practical application potential of current technology. Existing systems have made significant progress in terms of recognition accuracy, cross-domain adaptation, and real-time performance. However, they still face challenges, including insufficient cross-environment robustness, a limited ability to learn from fewer samples, and high system resource overhead. Future research should focus on adaptive and sample-less learning, edge computing, low cost, and privacy preservation to promote the wider application of CSI-based gesture recognition technology.

References

1. J. Su et al., A real-time cross-domain Wi-Fi-based gesture recognition system for digital twins. IEEE J. Sel. Areas Commun. 41, 3690-3701 (2023)
2. R. Gao et al., WiCGesture: Meta-motion-based continuous gesture recognition with Wi-Fi. IEEE Internet Things J. 11, 15087-15099 (2024)
3. Y. Chen, X. Huang, WiGNN: WiFi-based cross-domain gesture recognition inspired by dynamic topology structure. IEEE Wirel. Commun. 31, 249-256 (2024)
4. X. Ding et al., Robust gesture recognition method toward intelligent environment using Wi-Fi signals. Measurement 231, 114525 (2024)
5. S.-H. Jeong et al., UbiGest: Smartphone-based ubiquitous gesture recognition with Wi-Fi. IEEE Internet Things J. 12, 6475-6491 (2025)
6. H. Yan et al., Wi-SFDAGR: WiFi-based cross-domain gesture recognition via source-free domain adaptation. IEEE Internet Things J. (2025)
7. C. Cao et al., Real-time cross-domain gesture and user identification via COTS WiFi. IEEE Trans. Mob. Comput. 24, 5124-5137 (2025)
8. Z. Zhao et al., CrossFi: A cross domain Wi-Fi sensing framework based on Siamese network. IEEE Internet Things J. 12, 20138-20155 (2025)
9. S. Tulli et al., Hand gesture recognition: A contemporary overview of techniques. 2024 Int. Conf. Autom. Comput. 457-463 (2024)
10. N. Mohamed et al., A review of the hand gesture recognition system: Current progress and future directions. IEEE Access **9**, 157422-157436 (2021)