Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher

Journal of Theory and Practice in Humanities and Social Sciences, Volume 2, Issue2, 2025
https://www.woodyinternational.com/
10.5281/zenodo.15187784

Research and Implementation of an AI-Based Human Posture Recognition System

Feng Miao, Qiang Zhang, Jun Yan

Cangzhou Normal University, Cangzhou, Hebei, China

Abstract: ***This study aims to construct a human posture recognition system based on artificial intelligence, with the goal of enhancing the accuracy and real-time performance of posture recognition. The system adopts a modular design that encompasses data collection, preprocessing, feature extraction, posture recognition, and result evaluation. By integrating multi-source data through a fusion strategy that combines visual and inertial sensor data, the system utilizes deep learning algorithms to achieve precise posture recognition. Experimental results demonstrate that the system excels in both sports training and medical rehabilitation fields. It can provide personalized training recommendations for athletes and assist coaches in optimizing training programs. Additionally, it helps doctors and therapists accurately monitor patient recovery progress, thereby improving rehabilitation outcomes. In the future, continuous optimization of the system will be pursued, expanding its application domains and providing stronger technical support for the intelligent development of related industries.***

Keywords: Artificial Intelligence; Human Posture Recognition; Multi-Source Data Fusion; Deep Learning; Sports Training.

Cited as: Miao, F., Zhang, Q., & Yan, J. (2025). Research and Implementation of an AI-Based Human Posture Recognition System. *Journal of Theory and Practice in Humanities and Social Sciences*, *2*(2), 45–52. Retrieved from https://woodyinternational.com/index.php/jtphss/article/view/194

1. Introduction

1.1 Research Background and Significance

Against the backdrop of rapid contemporary technological advancements, human motion pose recognition technology has emerged as a hotspot in interdisciplinary research, with widespread applications in sports training, medical rehabilitation, virtual reality, and other fields. It plays a significant role in enhancing human quality of life and work efficiency. In the realm of sports training, precise pose recognition can provide athletes with personalized technical analysis and guidance, helping them optimize movements and elevate their competitive performance. In the field of medical rehabilitation, this technology can monitor patients' actions in real-time during recovery training, offering scientific basis for adjusting treatment plans and accelerating the rehabilitation process. The virtual reality domain, by leveraging pose recognition, enhances immersive experiences, expanding its application boundaries in gaming, education, and other scenarios (Rui Liang et al., 2025).

Traditional pose recognition methods are limited by their underlying technological principles and have several shortcomings. For instance, vision-based approaches are highly sensitive to environmental lighting conditions, with slight changes in illumination potentially leading to recognition errors. While sensor-based perception can acquire precise data, the equipment is complex and the data processing procedures are cumbersome, making it difficult to meet real-time requirements (Sujoy Karmakar et al., 2024). These limitations have severely restricted the widespread adoption and deeper application of human motion pose recognition technology.

In recent years, the swift development of artificial intelligence technology has provided an opportunity to break through the aforementioned challenges. The integrated application of cutting-edge technologies such as deep learning and computer vision has made it possible to achieve efficient, accurate, and intelligent human motion pose recognition in complex environments, carving out new avenues for research in this field(K. Aarthy & A. Alice Nithya, 2024).

1.2 Research Objectives and Questions

This study aims to construct a human motion pose recognition system based on artificial intelligence, with the goal of significantly improving the accuracy and real-time performance of pose recognition, thereby promoting the deepened application of this technology across multiple domains. Centering on this objective, this research delves into the exploration of the following key questions: How to effectively integrate multi-source data, combining visual information with sensor data to maximize the extraction of data value; How to accurately select and optimize deep learning algorithms to achieve optimal performance in complex pose recognition tasks; How to scientifically and reasonably optimize the system architecture to ensure efficient and collaborative operation of various modules, meeting the diverse needs of practical applications. Through systematic research and practice on the above issues, this study expects to provide innovative solutions and theoretical support for the development of human motion pose recognition technology.

2. Research Review

2.1 Traditional Human Motion Pose Recognition Methods

In the field of human motion pose recognition, traditional methods mainly rely on two technological approaches: visual perception and sensor perception, which have laid the foundation for the subsequent introduction of artificial intelligence technology.

In terms of visual perception, motion assessment methods based on two-dimensional videos hold an important position. For example, the convolutional pose machine technology analyzes each frame in the video to accurately identify human joint points, thereby achieving pose evaluation. The advantage of this method lies in its ability to capture relatively comprehensive motion information, especially suitable for single-person large-range motion scenarios. This method requires a high demand for computational resources, and in environments with multiple people or complex backgrounds, it is prone to joint point confusion, resulting in a decline in recognition accuracy (Amit Baghel et al., 2025).

Sensor perception is represented by Inertial Measurement Units (IMUs). By wearing sensors on different parts of the body, data such as acceleration and angular velocity can be obtained. Its advantage lies in the precision and high real-time performance of the data, which are not affected by environmental factors such as lighting. The wearing of sensor equipment may restrict the user's movements, and the data fusion processing is relatively complex, requiring sophisticated algorithms to ensure accurate integration of multi-source data (Youssef Nkizi & Ornwipa Thamsuwan, 2024).

2.2 Application of Artificial Intelligence in Pose Recognition

With the rapid development of artificial intelligence technology, deep learning algorithms have gradually become the core driving force in the field of human pose recognition. Among them, Convolutional Neural Networks (CNN) and their variants, as well as Recurrent Neural Networks (RNN), have performed exceptionally well in numerous research and applications (Xiongfeng Li et al., 2024).

CNN, with its unique convolutional and pooling layer structure, can automatically extract features from images and perform excellently in keypoint detection tasks. Taking OpenPose algorithm as an example, it utilizes CNN to extract features from images and then predict the locations of human joint points, achieving remarkable results in single-person pose estimation. When faced with complex backgrounds or scenarios with multiple people overlapping, its performance will be affected to some extent (Wei Zhao et al., 2024).

RNN is good at processing sequence data and performs outstandingly in action classification tasks. For example, Long Short-Term Memory networks (LSTM) can effectively model the inter-frame information in video sequences, thereby accurately recognizing continuous actions. But the training process of RNN is relatively complex and has high requirements for data preprocessing (Yang Min et al., 2021).

In recent years, models based on Transformer have begun to emerge. Utilizing the self-attention mechanism, they exhibit unique advantages in global feature extraction and long sequence modeling. On large-scale datasets, Transformer models show excellent recognition accuracy and generalization ability, providing new solutions for complex pose recognition tasks (Hong Zhao et al., 2025).

2.3 Multi-source Data Fusion Technology

Multi-source data fusion technology aims to integrate information from different sensors or data sources to improve the overall performance and reliability of the system, and it is an important development direction in the field of human motion pose recognition (Haoran Zhan et al., 2024).

Kalman filter, as a classic linear filtering algorithm, is widely used in sensor data fusion. By recursively estimating the state space model, it effectively fuses measurement data from different sensors, reduces noise influence, and improves the accuracy and reliability of the data. The Kalman filter has limitations in dealing with nonlinear systems and requires appropriate extensions or improvements (Mohammad Mahdi Rusta et al., 2024).

Particle filter is a nonlinear filtering algorithm based on the Monte Carlo method, capable of handling more complex system models. By introducing a large number of particles to represent the probability distribution of the system, particle filter can effectively fuse multi-source data, especially suitable for processing datasets with non-Gaussian noise or nonlinear relationships. But the computational complexity of particle filter is relatively high, requiring a balance between the number of particles and computational efficiency (Jiaying Lan et al., 2019).

Sensor information fusion algorithms are also constantly developing and innovating. By designing reasonable weight distribution strategies, data from different sensors can be weighted and fused to fully leverage the advantages of each sensor and compensate for the shortcomings of a single sensor (Jiale Qiao et al., 2019). Deep learning technology also provides new ideas and methods for multi-source data fusion. Neural network-based fusion models can automatically learn the complex relationships between data, achieving more efficient and accurate data (fusionKazuki HAYASHI & Daisuke TANAKA, 2023).

3. System Design and Implementation

3.1 System Architecture Design

The human motion pose recognition system constructed in this study adopts a modular design. The overall architecture, as shown in Figure 1, mainly consists of modules such as data acquisition, data preprocessing, feature extraction, pose recognition, and result evaluation. These modules achieve efficient data transmission and interaction through standardized data interfaces.

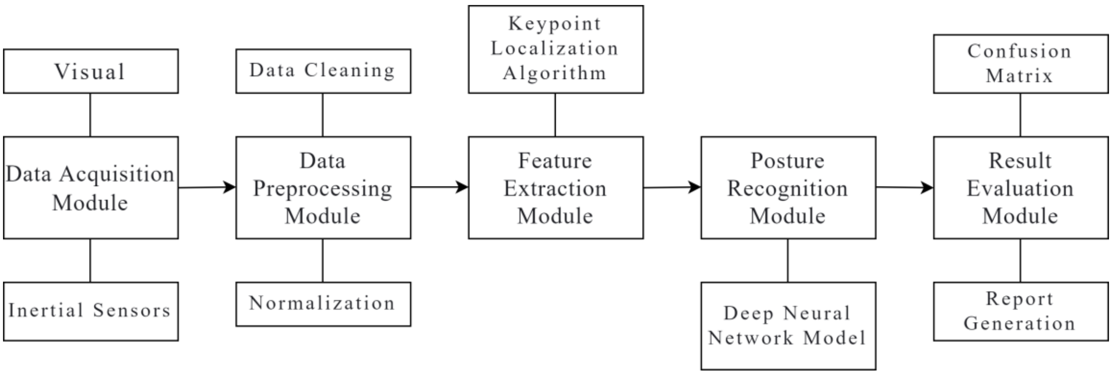

Figure 1: Human Motion Posture

The data acquisition module is responsible for obtaining raw data from multiple sensors, including visual information and inertial sensor data. Visual information is collected by high-definition cameras, which can capture the external manifestations of human movements. Inertial sensors accurately measure the acceleration and angular velocity of various parts of the human body, providing internal motion information. The data preprocessing module cleans and transforms the acquired raw data, removing noise and normalizing data formats to meet the requirements of subsequent processing. The feature extraction module utilizes a variety of artificial intelligence algorithms to extract key features such as joint positions and motion trajectories from the preprocessed data, providing a basis for pose recognition. The pose recognition module is the core of the system. It uses deep learning algorithms to analyze and classify the extracted features, achieving precise recognition of human motion poses. The result evaluation module quantitatively evaluates the recognition results and generates detailed reports, providing a basis for system optimization and improvement.

3.2 Data Acquisition and Preprocessing

Data Acquisition Equipment Selection and Arrangement

This system adopts a multi-source data acquisition strategy, comprehensively utilizing visual sensors and inertial sensors to fully capture human motion information. Visual sensors use high-resolution and high-frame-rate RGB-D cameras, which can simultaneously acquire color images and depth information. The cameras are fixed on tripods, and the height and angle are adjusted according to the application scenario to ensure the complete motion range of the human body is captured. Inertial sensors adopt multi-axis accelerometers and gyroscopes, which are worn on key parts of the human body such as the waist, knees, and wrists. The data is transmitted to the receiving device in real time via wireless transmission.

Data Preprocessing Methods

The raw data collected inevitably contains noise and interference, and needs to be preprocessed to improve data quality. For visual data, background modeling technology based on frame difference method is used to remove the static background and highlight the moving targets. A combination of Gaussian filtering and median filtering is used to further smooth the image and reduce the impact of noise. Inertial sensor data is processed by the Kalman filter algorithm to remove high-frequency noise and compensate for sensor drift and deviation. In addition, time synchronization processing is carried out for data from different sources to ensure consistency in the time dimension of multi-source data.

3.3 Feature Extraction and Selection

Artificial Intelligence-Based Feature Extraction Algorithms

In the feature extraction stage, this system comprehensively uses a variety of artificial intelligence algorithms to capture the key information of human motion poses accurately and comprehensively. The principal component analysis (PCA) algorithm is used for dimensionality reduction, extracting the main feature components in the data, reducing data redundancy while retaining key information (Kaige Ding et al., 2024). The sparse coding algorithm learns sparse representations from a large amount of data, highlighting significant features in the data and enhancing feature distinguishability. Deep learning algorithms, such as convolutional neural networks (CNN), automatically extract high-level semantic features from images, including human joint points and limb orientation.

Feature Selection and Optimization Strategies

The extracted features need to be further screened and optimized to improve the efficiency and accuracy of pose recognition. The feature selection method based on information gain evaluates the contribution of each feature to pose recognition and gives priority to features with high information gain. At the same time, regularization techniques are introduced to avoid overfitting problems and improve the generalization ability of the model.

3.4 Pose Recognition Algorithms

Improved OpenPose Algorithm

This system selects the improved OpenPose algorithm as one of the core algorithms for pose recognition. The traditional OpenPose algorithm performs well in single-person pose estimation but has certain limitations in complex scenarios (Sheng Bo et al., 2024). The improved OpenPose algorithm introduces an attention mechanism to enhance the model's focus on key parts and improve recognition accuracy. Specifically, an attention module is added in the network structure, enabling the network to automatically learn the importance weights between different joint points, thereby more accurately locating human poses.

Similarity Measurement-Based Regression Model

To improve the robustness of pose recognition, this system also introduces a regression model based on similarity measurement. This model constructs similarity measurements in the feature space and maps the input feature vector to the pose label space, achieving continuous estimation of poses. During training, the model learns the nonlinear relationship between features and poses, effectively handling the complexity of pose changes.

3.5 System Integration and Optimization

System Integration Method

Integrating each module into an efficient and stable human motion pose recognition system is an important part of this study. By designing a unified data transmission protocol and interface specifications, seamless docking between modules is ensured. A central control system is developed to schedule and coordinate each module, realizing automatic data flow and processing.

System Optimization Strategies

To meet real-time requirements, this system adopts a variety of optimization strategies. Parallel computing technology makes full use of the multi-core architecture of modern computers to parallelize computationally intensive tasks such as data processing and feature extraction, significantly improving system operation speed. Model compression technology prunes and quantizes deep learning models to reduce model size and lower computational complexity while maintaining high accuracy. In addition, parameter tuning is carried out on the system to determine the optimal parameter configuration of each module through experiments and data analysis, further improving system performance.

4. System Application and Case Analysis

4.1 System Application Fields

Sports Training

This system has significant application value in the field of sports training. By accurately identifying the athletes' movements, it provides coaches with detailed technical analysis data, including key information such as joint angles, limb positions, and motion trajectories. This data helps to formulate personalized training plans, optimize athletes' technical movements, and improve training effects. For example, in swimming training, the system can monitor the athletes' strokes and body posture in real time, promptly identify technical defects, and provide improvement suggestions. In sprint training, the system is able to capture the athletes' postures during the start, acceleration, and sprint phases, analyze key indicators such as stride length and frequency, and provide a scientific basis for improving performance.

Medical Rehabilitation

In the field of medical rehabilitation, this system can serve as an auxiliary tool for physical therapy. Doctors and therapists can use the system to accurately monitor patients' rehabilitation training processes and assess rehabilitation effects. The motion data provided by the system helps to adjust treatment plans in a timely manner and promote functional recovery of patients. For example, in the rehabilitation after knee arthroplasty, the system can monitor the range of motion of the patient's knee during walking and provide a scientific basis for rehabilitation training. In the field of neurological rehabilitation, such as the upper limb rehabilitation training of stroke patients, the system is able to capture the arm movement trajectory of patients in real time, analyze muscle strength and coordination, and provide data support for the formulation of personalized rehabilitation programs.

Virtual Reality

The requirements for immersive experiences in the field of virtual reality are becoming increasingly high. This system is able to provide natural and smooth human body motion posture recognition for virtual reality applications. In virtual games, virtual education, and other scenarios, users can interact with the virtual environment through natural actions, enhancing the user experience. For example, in virtual fitness games, the system can capture users' motion postures in real time, synchronize the virtual characters with users' actions, and enhance the realism and fun of the game. In virtual education scenarios, students can interact with virtual teaching content through natural gestures and body language, improving learning engagement and effectiveness.

Security Monitoring

This system also has important application potential in the field of security monitoring. By recognizing abnormal motion postures of personnel in the monitoring area in real time, such as falling, running, fighting, etc., alarms can be issued promptly to improve the efficiency and accuracy of security monitoring. In public places, important facilities, and other areas, the system is able to conduct intelligent analysis of personnel activities and achieve active prevention and rapid response to safety hazards. For example, in crowded places such as subway stations and shopping malls, the system can monitor abnormal behaviors of personnel in real time to ensure public safety.

Human-Computer Interaction

In the field of human-computer interaction, this system is able to realize more natural and intuitive human-computer interaction methods. Users can control devices or software directly through body movements without traditional input devices. For example, in smart home control, users can control the switching and adjustment of home appliances through gestures or body postures, improving the convenience of life. In the field of industrial automation, workers can work in coordination with robots through natural movements, improving production efficiency and safety.

4.2 Application Expansion and Technological Innovation

Multi-modal Data Fusion

This system supports multi-modal data fusion, combining visual information, inertial sensor data, depth data, and other data sources to achieve more comprehensive and accurate posture recognition. Through the complementarity and fusion of multi-modal data, the system can effectively address the deficiencies and limitations of a single data source and improve the recognition performance in complex environments. For example, in scenarios with insufficient lighting or severe occlusions, inertial sensor data can compensate for the lack of visual information, ensuring the continuity and accuracy of posture recognition.

Real-time Optimization

For application scenarios with high real-time requirements, this system adopts a series of optimization technologies. Through parallel computing, model compression, algorithm optimization, and other means, the processing speed of the system is improved to meet the needs of real-time interaction. For example, in virtual reality games, the system is able to capture and recognize users' motion postures at a high frame rate in real time, ensuring that the actions of the virtual characters are synchronized with the users and providing a smooth gaming experience.

Personalized Customization

This system can be customized according to the needs of different users. In sports training, it provides customized posture analysis and training suggestions for athletes of different projects and levels. In medical rehabilitation, it adjusts monitoring indicators and treatment plans according to the individual differences and rehabilitation progress of patients. In virtual reality applications, it optimizes posture recognition algorithms and interaction methods according to users' preferences and device configurations to meet diverse needs.

4.3 Application Prospects

With the continuous progress of technology and the expansion of application scenarios, the human body motion posture recognition system will show a broader application prospect. In intelligent sports, the system will be combined with more sports projects to provide athletes with full-scale technical support and training optimization schemes. In the field of smart healthcare, the system will be deeply integrated into rehabilitation therapy, remote medical treatment, and other links to improve the quality and efficiency of medical services. In the fields of virtual reality and augmented reality, posture recognition technology will promote the deep integration of virtual scenes and the real world and create a more realistic and natural interactive experience. At the same time, with the development of edge computing, 5G communication, and other technologies, the system will have stronger real-time performance and mobility, providing possibilities for applications in more fields.

5. Conclusion and Prospec

5.1 Summary of the Research

This study successfully constructed an artificial intelligence-based human motion posture recognition system aimed at improving the accuracy and real-time performance of posture recognition. The system adopts a modular design, covering data acquisition, preprocessing, feature extraction, posture recognition, and result evaluation. Each module realizes efficient collaboration through standardized interfaces. The innovation lies in the multi-source data fusion strategy, combining visual and inertial sensor data and using deep learning algorithms to achieve accurate posture recognition. Experimental results show that the system performs well in the fields of sports training and medical rehabilitation. It can provide personalized training suggestions for athletes and assist coaches in optimizing training programs. At the same time, it helps doctors and therapists accurately monitor the rehabilitation progress of patients and improve the rehabilitation effect. The system has efficient data processing and accurate posture recognition capabilities to meet the actual application needs.

5.2 Research Deficiencies and Prospects

Although this study has achieved significant results, there is still room for improvement. The limited amount of training data restricts the generalization ability of the model. It is necessary to expand the data set and incorporate more scene and individual data to improve the adaptability of the model. The high complexity of the algorithm makes it difficult to deploy on resource-constrained devices. Subsequent efforts should be made to optimize the algorithm and improve computational efficiency. In addition, the recognition accuracy of the system in complex environments needs to be further improved. In scenarios with strong light, occlusion, or multiple interferences, posture recognition may deviate. In the future, new deep learning architectures such as the combination of Transformer and convolutional neural networks can be explored to enhance the model's adaptability to complex environments. Fusing more types of data, such as physiological signals, is expected to further improve the accuracy and reliability of posture recognition. Through these improvements, the human motion posture recognition system will be continuously optimized and its application fields will be expanded to provide stronger technical support for the intelligent development of industries such as sports training, medical rehabilitation, and virtual reality.

Funding

This study was supported by the Self-financed Science and Technology Project of Cangzhou Science and Technology Bureau in 2022 (222105002).

References

[1] Rui Liang, Xing Luo, Bowen Zhao, Jiangxin Yang & Yanpeng Cao. (2025). Enhancing active steady-state non-line-of-sight recognition via a guided generative adversarial framework. Optics and Laser Technology, 183, 112269-112269.

[2] Sujoy Karmakar, Subhadeep Das, Subir Podder, Hriday Kumar Basak, Soumen Saha & Abhik Chatterjee. (2024). Synthesis, characterization and solvent effects on the electronic properties and non-linear optical activity of acridine-1, 8-diones-proline based fluorescent probe for visual recognition and selective detection of OH− ion. Journal of Molecular Liquids, 401, 124661-.

[3] K. Aarthy & A. Alice Nithya. (2024). Automated yoga pose recognition using enhanced chicken swarm optimization with deep learning. Multimedia Tools and Applications, 83(38), 1-23.

[4] Amit Baghel, Alok Kumar Singh Kushwaha & Roshan Singh. (2025). Automated Human Action Recognition with Improved Graph Convolutional Network-based Pose Estimation. International Journal of Pattern Recognition and Artificial Intelligence, (prepublish),

[5] Youssef Nkizi & Ornwipa Thamsuwan. (2024). Fall Risk Assessment in Active Elderly Through the Use of Inertial Measurement Units: Determining the Right Postural Balance Variables and Sensor Locations. Applied Sciences, 14(23), 11312-11312.

[6] Xiongfeng Li, Limin Zou & Haojie Li. (2024). Tai Chi Movement Recognition and Precise Intervention for the Elderly Based on Inertial Measurement Units and Temporal Convolutional Neural Networks. Sensors, 24(13), 4208-4208.

[7] Wei Zhao, Lei Wang, Yuanzhe Li, Xin Liu, Yiwen Zhang, Bingchen Yan & Hanze Li. (2024). A Multi-Scale and Multi-Stage Human Pose Recognition Method Based on Convolutional Neural Networks for Non-Wearable Ergonomic Evaluation. Processes, 12(11), 2419-2419.

[8] Yang Min, Zhang Yunong, Zhou Xuefeng & Hu Haifeng. (2021). Pose control of constrained redundant arm using recurrent neural networks and one-iteration computing algorithm. Applied Soft Computing Journal, 113(PB),

[9] Hong Zhao, Bojing Du, Yongju Jia & Hui Zhao. (2025). DanceFormer: Hybrid transformer model for real-time dance pose estimation and feedback. Alexandria Engineering Journal, 121, 66-76.

[10] Haoran Zhan, Jiange Kou, Yuanchao Cao, Qing Guo, Jiyu Zhang & Yan Shi. (2024). Human Gait phases recognition based on multi-source data fusion and BILSTM attention neural network. Measurement, 238, 115396-115396.

[11] Mohammad Mahdi Rusta, Seyyed Arash Haghpanah, Sajjad Taghvaei & Ramin Vatankhah. (2024). Adaptive neuro-fuzzy sliding mode control of the human upper limb during manual wheelchair propulsion: estimation of continuous joint movements using synergy-based extended Kalman filter. Neural Computing and Applications, 36(28), 17375-17416.

[12] Jiaying Lan, Guoheng Huang & Lianglun Cheng. (2019). Fast 3D Post Estimation of Human Based on Optical Flow and Particle Filter. Journal of Software, 437-448.

[13] Jiale Qiao, Jindong Zhang, Yuze Wang, Chenhui Yu & Sai Gao. (2019). A Multi-sensor Fusion System for Embedded Devices Considering by Sensor Reliability.(eds.)Proceedings of 2019 2nd International Conference on Mechanical Engineering, Industrial Materials and Industrial Electronics(MEIMIE 2019) (pp.88-91). Jilin University;

[14] Kazuki HAYASHI & Daisuke TANAKA. (2023). Effectiveness of Feature Extraction System for Multimodal Sensor Information Based on VRAE and Its Application to Object Recognition: Special Section on Deep Learning Technologies: Architecture, Optimization, Techniques, and Applications. IEICE Transactions on Information and Systems, E106.D(5), 833-835.

[15] Kaige Ding, Zhinan Zhao, Siyuan Ma, Yanqing Qiu, Tingting Lang & Ting Chen. (2024). Accelerating optimization of terahertz metasurface design using principal component analysis in conjunction with deep learning networks. Physical Communication, 66, 102452-102452.

[16] Sheng Bo, Chen Linfeng, Cheng Jian, Zhang Yanxin, Hua Zikai & Tao Jing. (2024). A markless 3D human motion data acquisition method based on the binocular stereo vision and lightweight open pose algorithm. Measurement, 225, 113908.