Journal of Engineering, Project, and Production Management
2026, 16(3), 2025-225

Optimized Sensor-Based Interaction System for Digital Media Games

Yu Ma
Associate Professor, School of Information Engineering, Xi'an Technology and Business College, Xi'an, 710200, China, E-mail: freelife_work@outlook.com

Project Management
Received October 13, 2025; revised November 26, 2025; accepted December 6, 2025
Available online April 8, 2026

Abstract: As technology advances at an accelerating pace, there is an increasing demand for games using digital media that offers advanced social interactions and an enhanced user experience. However, during 3D game development, issues such as inadequate capture of human body posture information continue to be a challenge. To address this, the study introduces a training method for a digital media game interaction system based on the dual-sensor Levenberg-Marquart (LM) algorithm to improve the system's computational capabilities. This approach also employs Composite Fields for Human Post Estimation (PIFPAF) to extract key features of human body movement points, boosting the system's ability to handle complex scenarios. Experimental results indicate that when the algorithm iterates between 200 and 250 times, the player's interaction loss rate ranges from 0.03% to 0.04%. The proposed method responds to all four types in under 0.250 seconds. In human-computer interaction, delays must remain below the human perception threshold to maintain immersion. Therefore, the system's ability to respond rapidly ensures that game interactions occur closer to real-time while effectively avoiding operational lag and guaranteeing a more natural and smooth interaction for its users. In contrast, the bidirectional long short-term memory training method is 0.350 seconds, and its perceived delay is more likely to disrupt the player's immersive state. These outcomes demonstrate that the proposed training method effectively captures human body information and facilitates Human-Computer Interaction feedback. This research contributes to the future development of 3D digital media games by supporting more advanced interaction commands.

Keywords: Human-Computer Interaction (HCI), Levenberg–Marquart (LM) algorithm, sensor technology, 3D game development, digital media systems.

DOI 10.32738/JEPPM-2025-225

---

1. Introduction

With the rapid growth of digital media games, player's demand for a more engaging experience is increasing (Guo and Lo, 2024). Traditional game interaction methods include controller-based and touchscreen interactions. Although these methods are convenient and efficient, they often have high error rates and lack precision (Onyejelem and Aondover, 2024). Consequently, many experts have researched digital media game interaction systems. Some have developed more efficient game engines to improve graphics rendering. For instance, Unreal Engine 5's Nanite virtual micro-polygon geometry technology enables real-time streaming and scaling of artistic works, maintaining stable frame rates without significant distortion (Saputri and Indriayu, 2024). While these technologies provide high-quality visual experiences, they struggle to accurately capture multidimensional information such as player movements and positions (Lyu, 2024). To address this, many game interaction designers use the LM algorithm to optimize multidimensional data. Furthermore, the LM algorithm calibrates motion capture devices and adjusts parameters such as animation speed and amplitude in real-time based on player input (Fischer et al., 2024). However, the LM algorithm has high computational complexity and can become stuck in local optima during scene rendering (Uwimana et al., 2025). To overcome issues like biased human body capture and limited behavior prediction. This study proposes a training method based on a dual-sensor LM algorithm. Unlike previous LM-based optimizations that primarily focused on parameter calibration or single-sensor data, this research integrates the Dragonfly Algorithm (DA) for global optimization and employs dual-sensor fusion for multimodal data (e.g., sound and motion), thereby enhancing the system's robustness in complex gaming scenarios. Moreover, the proposed framework incorporates the Composite Fields for Human Pose Estimation (PIFPAF) algorithm for fine-grained body posture capture, which surpasses traditional skeleton-based methods in describing complex structural information. This integrated approach combines bio-inspired optimization, multimodal sensing, and composite field-based pose estimation, representing a novel contribution to the field of Artificial Intelligence (AI), which assists with interactive gaming, particularly in enhancing real-time responsiveness and multiplayer interaction accuracy.

2. Related Works

To meet the immersive demands of digital game interaction with humans, many scholars have researched digital media interaction technologies. Zargham et al. (2025) conducted in-depth interviews with experts in dialog user interfaces and game user studies, aiming to improve single-player video and voice interaction capabilities. The results show that voice interaction in games promotes greater immersion, engagement, and entertainment. Addressing the requirements of urban space gamification, Vilar et al. (2025) proposed a location-based mixed reality solution. This solution used interactive maps and augmented reality technology to support the development of gamified applications. The results showed that this approach improved the interactive narrative capabilities of urban spaces. Hare et al. (2025) applied multi-agent reinforcement learning to enhanced user interaction methods in AI-integrated systems. The experimental results showed that this method provides positive feedback throughout the process of automated tutoring in gamified virtual environments. Aiming to achieve deep integration of the Internet and education, Yang and Bao (2025) designed a remote language learning platform based on Human-Computer Interaction (HCI) technology. The experimental results showed that the system performed well for large-scale users, with fast message transmission rates and significant advantages in remote online interaction effects. Addressing issues such as incomplete video feature extraction in human-computer interaction, Sun and Kwak (2024) proposed a multimedia human-computer interaction method for animated works based on convolutional networks. The experimental results showed that this approach was practical in multimedia courseware, providing a reference for modern HCI research. With the upgrade of HCI technology, demands for digital media game interaction have increased. Not only is there a need for real-time capturing of player game dynamics, but also for improved visual effects and realism in games. Therefore, researchers have focused on the LM algorithm. To extend the lifespan of wireless sensor networks, Revanesh et al. (2024) proposed an improved LM neural network utilizing artificial neural networks for energy efficiency and anomaly detection in wireless sensors. The experimental results showed that this algorithm could be used for intrusion detection systems and performed well in identifying anomalies. Furthermore, to address issues of instability in integrated electric power quality regulators, Srilakshmi et al. (2024) used the LM algorithm for backpropagation training of an artificial neural network control. The experimental results show the proposed method significantly reduced harmonic distortion to 3.61%, 3.48%, 3.48%, and 4.51%, and unified the power factor. To solve the problem of battery energy consumption in wireless sensor network nodes, Hakim et al. (2024) proposed an LM-artificial neural network self-organizing network model. The experimental results showed that the model could adaptively adjust parameters (Hakim et al., 2024). To accurately predict solar irradiance, Yuzer et al. (2024) used the LM algorithm and transfer functions to train artificial neural networks. The experimental results showed that this training algorithm could provide accurate predictions even without direct irradiation measurements. Zhang et al. (2024) In order to estimate the health status of power systems, Zhang et al. (2024) proposed a method based on the LM backpropagation neural networks and interleaved step-up converter systems to extend the power wiring lifespan. The experimental results showed that this method could extend the system's lifespan and hold significant implications for predictive maintenance in industrial systems. Scholars both domestically and internationally have conducted detailed research on digital media interaction systems, achieving notable success in artificial interaction technologies. However, there is limited research combining digital media game interaction systems with the LM algorithm. Therefore, this study designs a training method for digital media game interaction systems based on a dual-sensor LM algorithm to enhance the gaming experience. The aim is to offer a more convenient and efficient training approach for intelligent interaction systems, meeting real-time operation requirements for different players.

3. Design a Digital Media Game Interaction System Based on a Sensor and an LM Algorithm

3.1. Improvement of the LM algorithm Based on Dual-Sensors

With the development of the internet, HCI technology has become a bridge connecting humans and intelligent systems, significantly transforming how people interact with technological devices (Yokoi and Nakayachi, 2025; Gil et al., 2024). However, there are still some technical limitations in the advancement of HCI technology, such as inaccurate gesture recognition and challenging operation. Therefore, this study proposes an LM algorithm based on dual sensors to address issues such as inaccurate data collection in HCI systems and to precisely adjust parameters for computer vision tasks. The LM algorithm offers faster convergence and improved recognition of image objects (Liao et al., 2025). The damping coefficient adjustment is shown in Eq. (1).

$$\alpha = \alpha_d + \frac{1}{1 + \exp\left(-\frac{r(p)_i - r(p)_{i-1}}{r(p)_{i-1}} \right)} \delta \tag{1}$$

In Eq. (1), $r(p)$ and α represent the residual vector and fixed adjustment factor, while α_d and δ represent the lower limit and adjustment factor range, respectively. The LM gradient descent for finding the optimal solution is shown in Eq. (2).

$$x_{n+1} = x_n - \alpha \cdot \nabla f(x_n) \tag{2}$$

In Eq.(2), α and $\nabla f(x_n)$ represent the learning rate and gradient, while n and $f(x_n)$ represents the step size and function, and x_n represents the solution. To further capture human body movements in the HCI system, the study also improves the LM algorithm by using the DA, which enhances the algorithm's convergence. DA has advantages such as strong global search ability and good robustness (Avcı and Akgül, 2024). The process of the DA-improved LM algorithm is shown in Fig. 1.

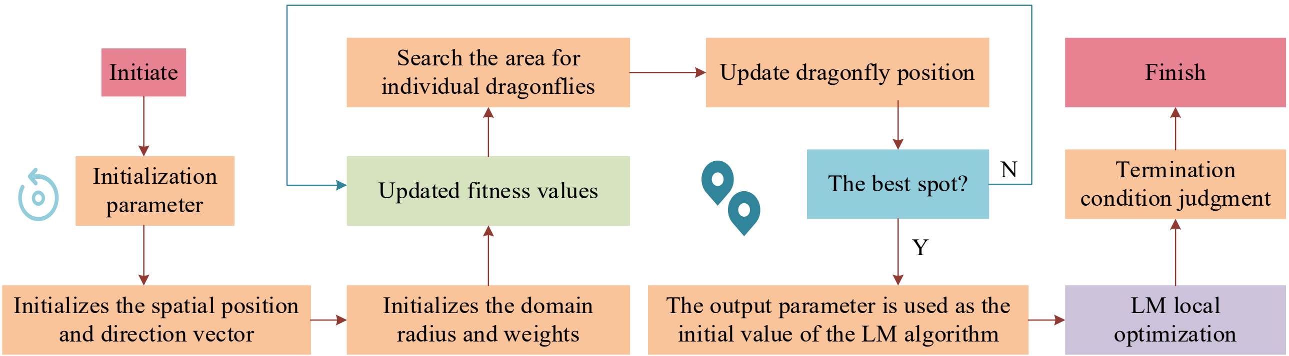


Fig. 1. Flowchart of the LM algorithm improved based on DA

As shown in Fig. 1, the DA-improved LM algorithm process begins by initializing parameters, setting the maximum iteration count and population size. Then, the initial position and direction vectors in the initialization space are assigned random values between -100 and 100. Next, the neighborhood radius and separation weight are initialized. The fitness values are then calculated and updated, searching for dragonfly individuals in the neighborhood and adjusting their behavior, position, and timing. After that, it is determined whether the current calculated position is optimal. If it meets the criteria, the parameters are output as initial values for the LM algorithm. The process then advances to the LM local optimization phase, where parameters are iteratively updated, and error reduction values are measured. Finally, convergence conditions are checked, and if satisfied, the process terminates. The constraints of the improved LM are shown in Eq. (3).

$$E(\theta,d) = {}_{\omega IK} E_{IK}(\theta,d) + {}_{\omega s} E_s(\theta,d) + {}_{\omega d} E_d(\theta,d) \tag{3}$$

In Eq. (3), $E_{IK}(\theta,d)$ represents the inverse kinematic constraint, $E_s(\theta,d)$ represents the human data smoothing constraint, $E_d(\theta,d)$ represents the human body capture depth constraint, and θ represents the Euler angles. Although the DA-improved LM algorithm can effectively capture human body data, it cannot capture signals like sound and vision. Therefore, based on the DA-improved LM algorithm, the study uses dual sensors to accurately calculate information such as sound sources. Dual sensors can collaborate with the HCI system to obtain more accurate position and posture information. The specific structure for sound source localization using dual sensors is shown in Fig. 2.

As shown in Fig. 2, the specific structure for sound source localization using dual sensors first gathers signals from sound sources. The data collected from accelerometers, gyroscopes, pressure sensors, and similar devices reflect real-time information like object position, speed, and posture. This data is then sent to the recognition sound card for sound signal processing. Afterward, the processed data is transmitted to the analysis platform. Next, the data is sensor-driven by the HCI system, which uses the Global Positioning System (GPS) and Inertial Measurement Unit (IMU) modules to accurately measure object acceleration, magnetic field, and other parameters, thereby determining the object's motion state. Finally, the positioning results are output in either two-dimensional or three-dimensional space, with label information. The sound source position calculation is shown in Eq. (4).

$$x_{m1}(n) = x_{m1}(n)^* \min(n) \tag{4}$$

In Eq. (4), m_1 represents the coordinates of sensor one, n and $\min(n)$ represent the corresponding coordinates of the sound source and the minimum distance coordinate, respectively. The cross-correlation function between the dual sensors is expressed in Eq. (5).

$$R_{m1m2}(\tau) = h(\tau)^{*} R_{m1m2}(\tau) \tag{5}$$

In Eq. (5), $R_{m1m2}(\tau)$ and represent the cross-correlation function and interpolation filter between the dual-sensors, and $R_{m1m2}(\tau)$ represents the peak value between the dual sensors. The dual-sensor error correction deviation is shown in Eq. (6).

$$r = 1-(x-o_x)^2 s_x^2 + (y-o_y)^2 s_y^2 + (z-o_z)^2 s_z^2 \tag{6}$$

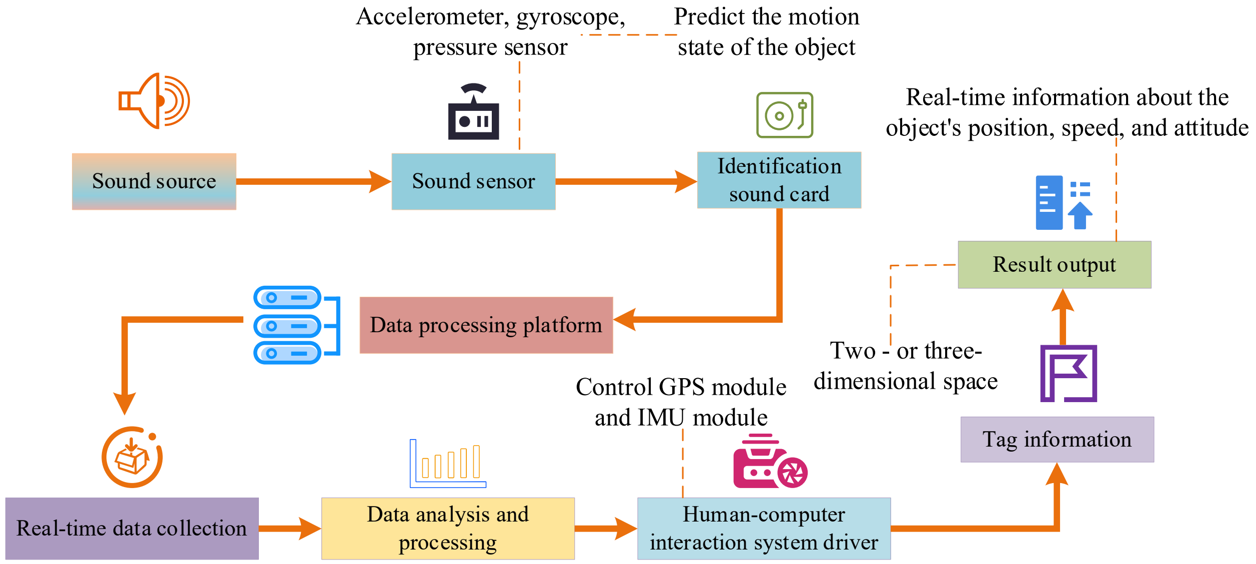


Fig.2. Flowchart of sound source localization using dual sensors

In Eq.(6), (x,y,z) represents the measured position coordinates of the sensor, s_x^2, s_y^2, and s_z^2 represent the variance-related quantities, and (o_x,o_y,o_z) represents the actual, accurate position coordinates. In summary, dual sensors can accurately measure and transmit the position of different objects. The specific framework for the LM algorithm-based HCI system with dual sensors is shown in Fig. 3.

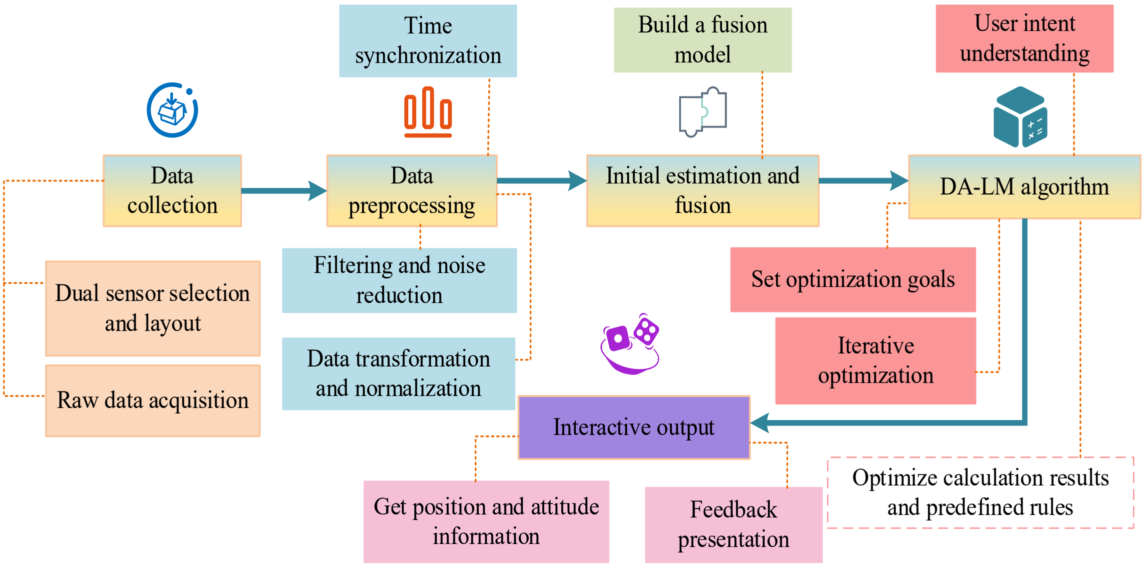


Fig. 3. HCI system framework based on dual sensor improvement

As shown in Fig. 3, the LM algorithm-based HCI system with dual sensors begins with sensor data collection. Visual, audio sensors, and other devices gather real-time data from the user to ensure effective processing later. Then, the dual-sensor data and other inputs are time-matched and converted into spatial coordinates, with noise filtered out and features normalized. The DA-improved LM algorithm model is then used to optimize the integrated data, continuously refining the robot's position and posture estimate and inferring user intentions. Finally, the HCI output is processed, delivering feedback through image displays and voice prompts.

2.2. Optimization of the Digital Media Game Interaction System Training Method Combining LM

Although the LM algorithm with dual sensors can accurately capture posture information for HCI systems, there are still some limitations in digital media game interaction systems. Today, digital media game interaction systems can effectively transmit and exchange information between players and games. However, issues like weak image recognition and technical instability persist in these systems. Therefore, based on the DA-improved LM algorithm, this study proposes a game interaction method using PIFPAF to obtain pixel coordinates of image key points. The PIFPAF algorithm can not only precisely locate body parts but also describe complex structural information more accurately. The structure of the PIFPAF algorithm is shown in Fig. 4.

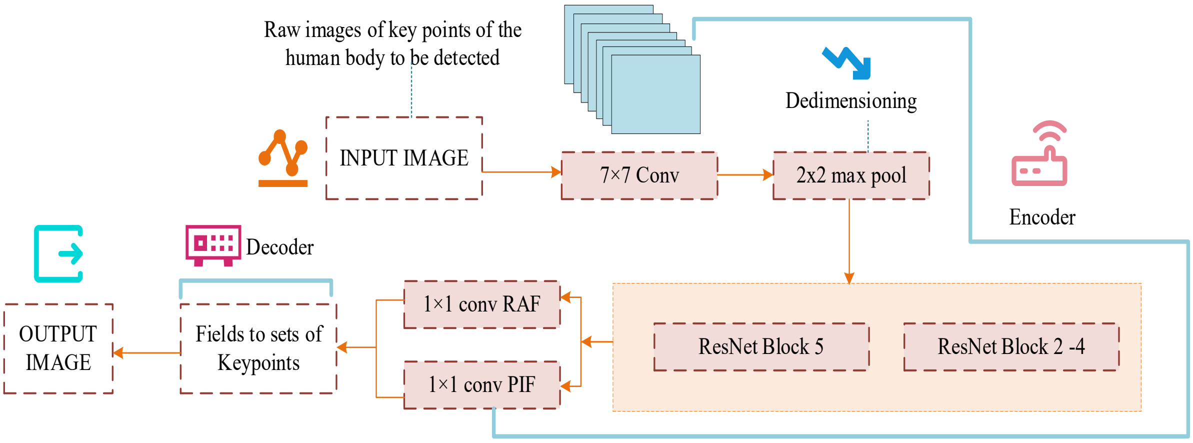


Fig. 4. PIFPAF algorithm structure diagram

As shown in Fig. 4, the PIFPAF algorithm structure consists of the input layer, encoder, middle layer, decoder, and output layer. First, the input layer inputs the original image data of the human key points to be detected. Second, the data enters the encoding layer, where a 7×7 convolution is used to extract features from the input image, followed by a 2×2 max-pooling operation to reduce the data dimension. The network layers are deepened using residual network modules, further refining the image features. The 1×1 Part Association Field (PAF) and Part Intensity Field (PIF) generated by the residual network are then decoded and transformed into a set of human key points. Finally, the detected game human key points are output. The PIF branch network output calculation is shown in Eq. (7).

$$P^{ij}=\left\{p_c^{ij},p_x^{ij},p_y^{ij},p_b^{ij},p_\sigma^{ij}\right\} \tag{7}$$

In Eq. (7), i and j represent the output heat map coordinates, while the pixel confidence map and the components of the nearest key points in the x and y directions are represented by p_c^{ij}, p_x^{ij}, and p_y^{ij}, respectively, and p_b^{ij} and p_σ^{ij} represent the loss function correction parameters and Gaussian kernel smoothing parameters. The RAF branch key point prediction is shown in Eq. (8).

$$A^{ij}=\left\{a_c^{ij},a_{x1}^{ij},a_{y1}^{ij},a_{b1}^{ij},a_{x2}^{ij},a_{y2}^{ij},a_{b2}^{ij}\right\} \tag{8}$$

In Eq. (8), a_{x1}^{ij} represents the component of the pixel in the first offset vector x direction of the limb, a_{y1}^{ij} represents the component in the y direction, and a_c^{ij}, a_{b1}^{ij}, and a_{b2}^{ij} represent the confidence map and correction parameters. The loss function calculation is shown in Eq. (9).

$$L=\frac{1}{N}\sum_{i=1}^{N}-\left[y_i\cdot\log\left(p_i\right)+\left(1-y_i\right)\cdot\log\left(1-p_i\right)\right] \tag{9}$$

In Eq. (9), N and y_i represent the sample quantity and sample labels, while p_i represents the correct class probability predicted by the sample. To provide users with a better experience in the digital game interaction system, the study combines Virtual Reality (VR) technology with the PIFPAF game interaction method to enhance the player's ability to interact with objects in the simulated environment. VR has advantages such as multi-sensory engagement and high safety. The process of digital media game interaction using VR is shown in Fig. 5.

As shown in Fig. 5, the process of digital media game interaction using VR is divided into image collection, AI, data interaction modules, game running platform, and playback rendering effects. First, VR cameras are used to collect images and extract action data for players in the digital media game interaction system. Second, the collected data enters the media interaction module, where it is encapsulated and sent to the game running platform in an analyzable format. Third, the rendered game screen is displayed in real-time to the player who watches and interacts. The specific expression of the input variables after fuzzy processing is shown in Eq. (10).

$$\mu = e^{\frac{\left(x_q - c_{qw}\right)^2}{\sigma_{qw}^2}} \tag{10}$$

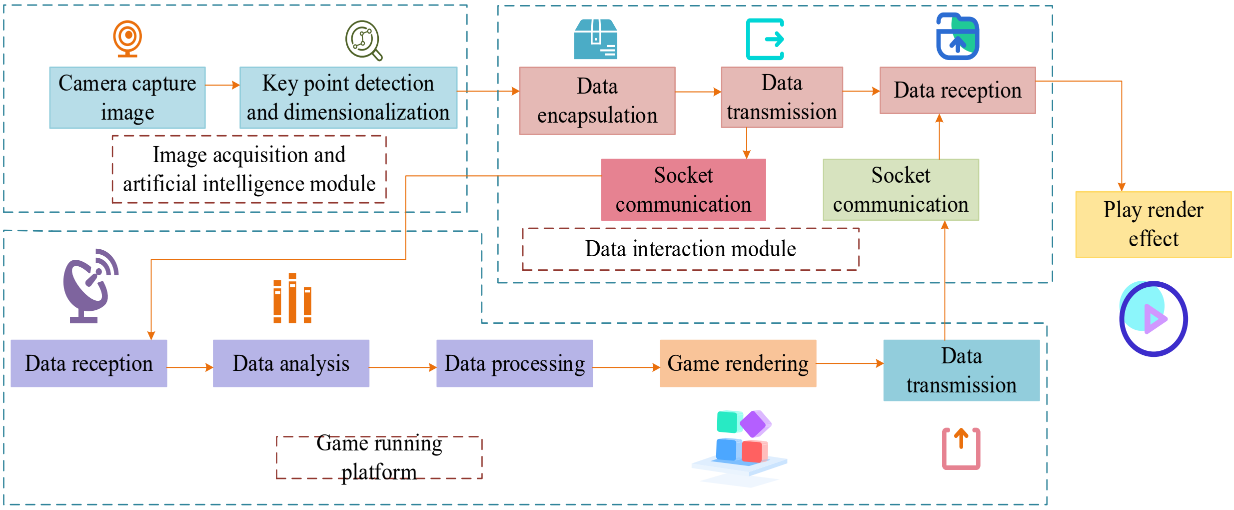


Fig. 5. The process of using VR to interact with digital media games

In Eq. (10), c_{qw} and σ_{qw} represent the Gaussian function center value and width value, while e and x_q represent the natural logarithm and corresponding nodes, and q and w represent input layer nodes. The description of the nodes between is shown in Eq. (11).

$$O_{q,w}^{(r)} = \frac{1}{1 + \exp\left(-n_{q,w}^{(r)}\right)} \tag{11}$$

In Eq. (11), $O_{q,w}^{(r)}$ and $n_{q,w}^{(r)}$ represent the output value and input value of the nodes. The digital media game interaction system formalizes the calculation of user queries as shown in Eq.(12).

$$F\left(A_R^k \middle| C_H^k, E_H^k\right) \to C_R^k \tag{12}$$

In Eq. (12), A_R^k and C_H^k represent the candidate reply set and player dialogue content, while E_H^k represents the player's emotional analysis. In summary, the digital media game interaction method using PIFPAF and VR technology can better capture the player's body language information. The structure of the digital media game interaction system training method based on the dual-sensor LM algorithm is shown in Fig. 6.

As shown in Fig. 6, the training process for the digital media game interaction system based on the dual-sensor LM algorithm is divided into five modules: data collection, data preprocessing, an improved LM algorithm, interaction strategy, and evaluation and optimization. First, the data collection module uses dual sensors to capture body language and other information. Second, the collected data is sent to the data preprocessing module for noise filtering, data normalization, and feature extraction. Next, the improved LM algorithm processes the sequence data and analyzes player behavior. The trained LM model is integrated into the game logic to adjust interactions in real-time. Finally, a feedback mechanism provides game status information to both the player and the system, enabling timely system optimization. This method is highly suitable for interactive gaming scenarios requiring high real-time responsiveness.

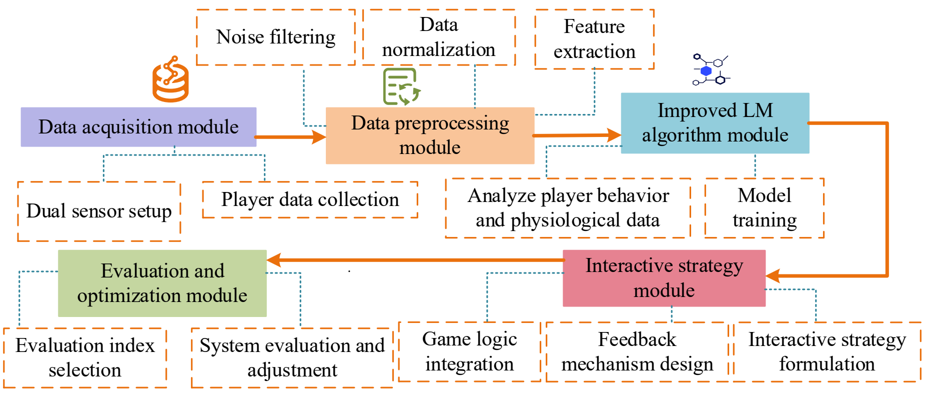
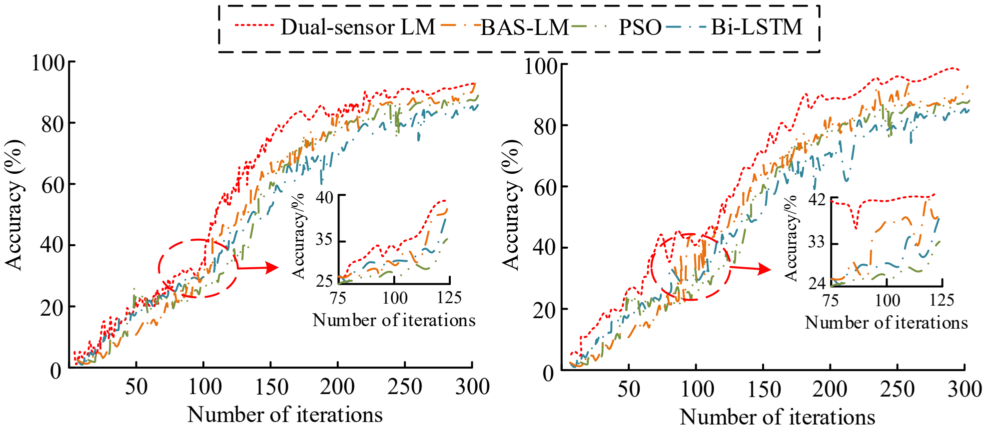


Fig. 6. Structure of the proposed digital media game interactive system training method

4. Performance Evaluation of the Digital Media Game Interaction System Training Method

4.1. Performance Analysis of the LM Algorithm based on Dual Sensors

To verify the superior performance of the dual-sensor LM algorithm, the study compared it with the Beetle Antennae Search-LM (BAS-LM) algorithm, the Particle Swarm Optimization (PSO) algorithm, and the Bidirectional Long-Short Term Memory (Bi-LSTM) algorithm. The experimental environment consisted of an AMD Ryzen 7 4800H with Radeon Graphics 2.90 GHz CPU, Windows 10 operating system, 16GB memory, and a Structured Light depth camera for image collection, with human body image data captured using the Kinect V2 sensor. To ensure the authenticity and reliability of the experiments, the study used the OpenIGDB dataset and the MovieLens-Games dataset for testing and training. The study conducted player posture estimation accuracy tests with the four algorithms: dual-sensor LM, BAS-LM, PSO, and Bi-LSTM, and the results are shown in Fig. 7.



(a) Test results of attitude prediction accuracy in OpenIGDB dataset

(b) Test results of attitude prediction accuracy in MovieLens-Games dataset

Fig. 7. Accuracy results of posture estimation in different data sets

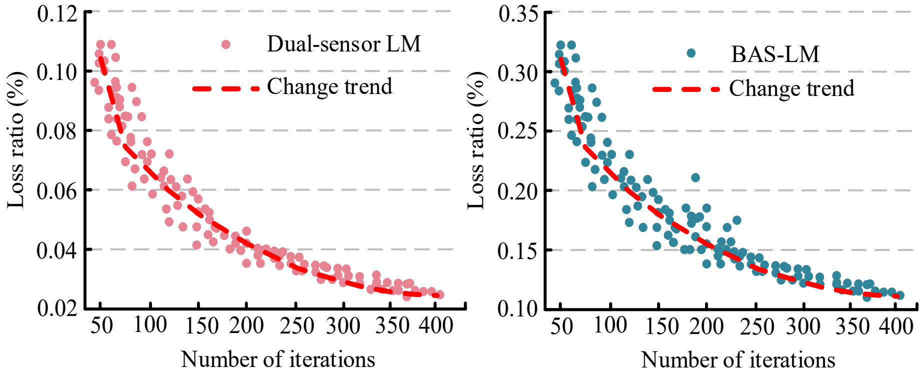
As shown in Fig. 7(a), when tested on the OpenIGDB dataset, the dual-sensor LM algorithm's posture estimation accuracy steadily improved after 125 iterations, reaching a peak of 98.76%. As shown in Fig. 7(b), during testing on the MovieLens-Games dataset, after 100 iterations, the dual-sensor LM algorithm achieved a player posture estimation accuracy of 42.89%, while the BAS-LM algorithm reached 38.97%. After 200 iterations, the dual-sensor LM algorithm accuracy increased to 87.86%. In conclusion, the dual-sensor LM algorithm can reliably estimate player posture with a certain level of stability. To validate the fitting performance of the dual-sensor LM algorithm, the study compared it with BAS-LM, PSO, and Bi-LSTM in terms of player interaction loss rate, and the results are shown in Fig. 8.

As shown in Fig. 8(a), when the dual-sensor LM algorithm is iterated 200 to 250 times, the player interaction loss rate can be controlled at an extremely low level of 0.03% to 0.04%. This extremely low loss rate means that the system hardly misses or delays when recognizing player action instructions, thereby significantly enhancing the game's response speed and operational smoothness, providing players with a continuous and stable immersive interactive experience. After 250 iterations, the loss curve gradually stabilized, further indicating that the system has good robustness and stability in long-term operation. In Fig. 8(b), the BAS-LM algorithm showed large fluctuations in the player interaction loss rate between 150 and 200 iterations. Fig. 8(c) indicates that the PSO algorithm had a player interaction loss rate between 0.12% and 0.19% during iterations from 200 to 300. As depicted in Fig. 8(d), the Bi-LSTM algorithm achieved a player interaction loss rate of 0.07% after 400 iterations. In summary, the dual-sensor LM algorithm can accurately respond to different player interactions and demonstrate good fitting performance. The study also evaluated the action capture F1-score of the four algorithms, with the results provided in. Fig. 9.

As shown in Fig. 9(a), when tested on the OpenIGDB dataset, the dual-sensor LM algorithm achieved the highest action capture of an F1-score of 0.97. During iterations from 150 to 200 times, the F1-score ranged between 0.56 and 0.72 times. After 200 iterations, the BAS-LM algorithm achieved an F1-score of 0.58. As shown in Fig. 9(b), during testing on the MovieLens-Games dataset, the dual-sensor LM algorithm's F1-score ranged from 0.37 to 0.59 between 100 and 150 iterations. After 200 iterations, the F1-score gradually stabilized, reaching a maximum of 0.98. In conclusion, the dual-sensor LM algorithm can accurately capture player actions, improving the gaming experience.

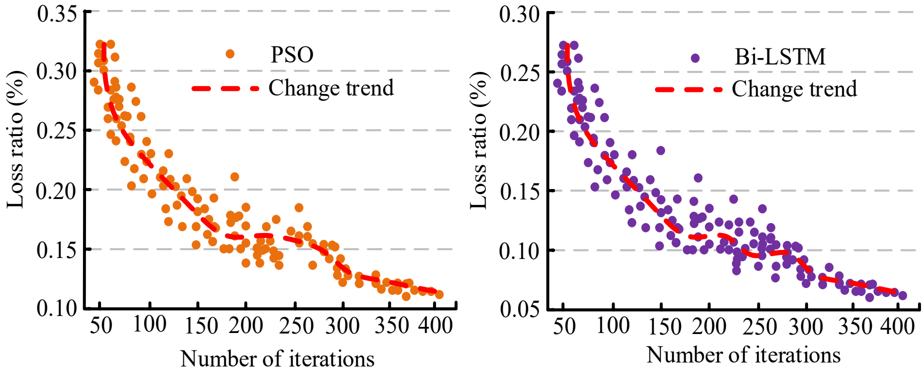
4.2. Effect Verification of Improved Digital Media Game Interaction System Training Method

After verifying the performance of the dual-sensor LM algorithm, the study further analyzed the proposed training method for the digital media game interaction system based on the dual-sensor LM algorithm by comparing it with the BAS-LM, PSO, and Bi-LSTM-based methods. The experimental platform consisted of a 13700k processor paired with a 4080 GPU, with the system running Ubuntu 20.04 LTS and using an NVIDIA GTX 980 graphics card. The UMass Game and Steam Video Games datasets were used, both of which are suitable for analyzing game-related text semantics and emotional tendencies. The four digital media game interaction system training methods were tested with online and offline platform question-answering, and the results are shown in Fig. 10.



(a) Test results of player interaction loss rate for dual-sensor LM algorithm

(b) Test results of player interaction loss rate for BAS-LM algorithm



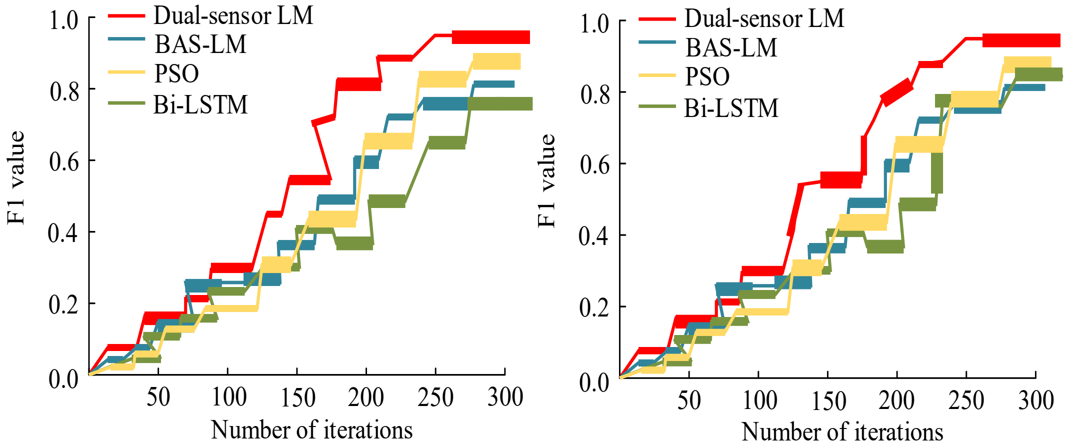
(c) Test results of player interaction loss rate for PSO algorithm

(d) Test results of player interaction loss rate for Bi-LSTM algorithm

Fig. 8. Test results of different algorithms on player interaction loss rate

As shown in Fig. 10(a), when conducting interactive tests with different players on the online platform, the dual-sensor LM digital media game interaction system training method achieved an average precision of 94.78%. As shown in Fig. 10(b), during intelligent question answering on the offline platform, the dual-sensor LM method reached an average precision of 87.36% after 20 interactions. In conclusion, the dual-sensor LM digital media game interaction system training method performed well on both online and offline platforms for player dialogue services and can meet various game service needs. Additionally, the study tested the interface gesture control execution time of the four digital media game interaction system training methods, and the results are shown in. Table 1.

As shown in Table 1, on the UMass Game dataset, the response time of the dual-sensor LM method for all four types of instructions is less than 0.250 seconds, and the buffer time does not exceed 0.200 seconds. The short buffer time means that the waiting process for the system to receive instructions is almost imperceptible to users, while the fast response time directly determines the immediacy of the feedback from the player's issuance of instructions to the game screen. These two indicators work together to ensure a high-quality user experience from two dimensions: system stability and operational smoothness, allowing players to fully immerse themselves in the game plot without being disturbed by technical delays. In conclusion, the training method of the dual-sensor LM digital media game interaction system can respond quickly and stably to various gesture game commands in the game interaction system, providing users with a seamless and natural gaming experience. The study also tested the facial expression prediction of players using the dual-sensor LM digital media game interaction system training method, and the results are shown in Fig. 11.

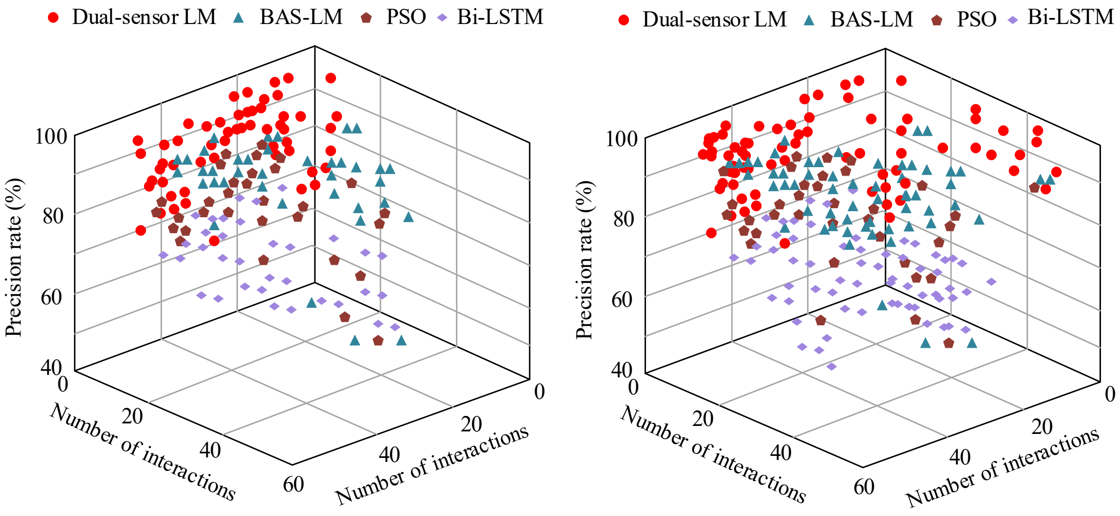


(a) Test results of F1-score for different algorithms in the OpenIGDB dataset

(b) Test results of F1-score for different algorithms in the MovieLens-Games dataset

Fig. 9. Test results of F1-scores in different datasets

As shown in Fig. 11, at 12s, the dual-sensor LM method predicted 52 different player facial expressions, which matched the correct number of expressions. Before 24s, there were significant errors in the predicted facial expression count, and the prediction curve fluctuated greatly. Between 25s and 27s, the prediction varied between 50 and 150, and at 27s, the predicted facial expression count was 75, close to the actual count. In conclusion, the research method is highly applicable to interactive game scenarios that require high real-time response requirements. The improvement in the performance of this system directly translates to a better gaming experience: higher pose estimation accuracy makes the movements of virtual characters more realistic and natural, thereby enhancing the player's sense of immersion. A lower interaction loss rate ensures accurate execution of operational intentions, enhancing the player's control confidence and participation. For developers, the stable and efficient interaction modules provided by this system can reduce the debugging time on the underlying motion capture and recognition and improve the development efficiency. In terms of application expansion, this system's solution is not limited to entertainment VR games. It can also be directly applied to scenarios that require high-precision human motion analysis, such as the medical rehabilitation field (such as the assessment and training of patient movement standardization) and the sports industry (such as the simulation and optimization of athlete technical movements).



(a) Online gaming platform question-answering

(a) Online gaming platform question-answering

Fig. 10. Interaction results between online and offline game platforms

Table 1. Results of gesture control execution time on the interface side

Data set	Order	Dual-sensor LM		BAS-LM		PSO		Bi-LSTM	
		Buffer time	Response time	Buffer time	Response time	Buffer time	Response time	Buffer time	Response time
UMass Game	1	0.025	0.143	0.037	0.169	0.058	0.198	0.061	0.184
	2	0.048	0.174	0.058	0.236	0.089	0.214	0.097	0.194
	3	0.146	0.178	0.213	0.289	0.201	0.236	0.223	0.298
	4	0.153	0.213	0.203	0.369	0.213	0.285	0.258	0.312
Steam Video Games	1	0.072	0.124	0.174	0.203	0.189	0.236	0.213	0.274
	2	0.126	0.189	0.198	0.224	0.201	0.259	0.197	0.203
	3	0.089	0.162	0.093	0.189	0.123	0.187	0.103	0.145
	4	0.063	0.103	0.089	0.174	0.176	0.189	0.072	0.113

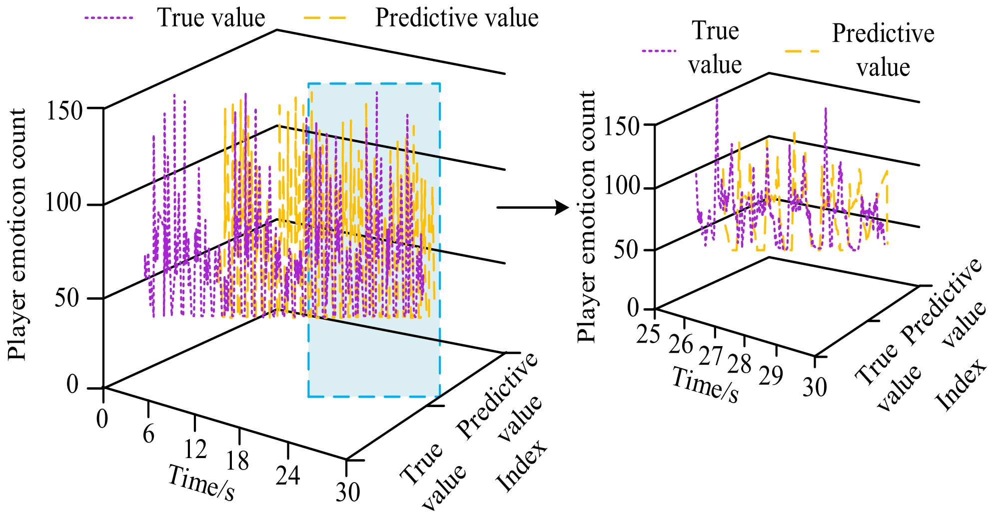


Fig. 11. Prediction results of the player's face using the dual sensor LM method

5. Conclusion

The study proposed a digital media game interaction system training method based on the dual-sensor LM algorithm to address the research gap in integrating global optimization, multimodal sensing, and composite human pose estimation within a unified gaming interaction framework. This approach effectively tackles issues such as limited motion capture capability and low operational efficiency prevalent in existing digital media game interaction systems. During the development of this method, the DA optimization technique was used to boost the system's global search capability, and the PIFPAF algorithm was integrated to extract key point information from composite structures, thereby enhancing the system's interaction capability. The experimental results show that the method proposed in this paper performs well in key performance indicators. For instance, the extremely low interaction loss rate (0.03%-0.04%) and millisecond-level response delay (<0.250s) not only reflect the algorithm's efficiency, but more importantly, they directly translate into the system's outstanding stability and smooth user experience. High-precision pose estimation ensures the faithful reproduction of player's movements by virtual characters, thereby essentially enhancing the immersion and interactive authenticity of 3D digital media games. This research not only enhances the technical performance of the interaction system through algorithm integration but, more importantly, its low latency and high precision features directly bring about a leap in quality, providing users with a deeper sense of immersion and a higher willingness to participate. The constructed system framework has the potential to expand into multi-user online scenarios and professional fields such as VR games, rehabilitation training, and sports simulation, providing a feasible solution for the application of digital media interaction technology in a wider range of industrial scenarios. Furthermore, an empirical analysis of the constructed digital media game interaction system training method revealed that, during interaction tests, the dual-sensor LM method achieved an average precision of 94.78%, while the BAS-LM, PSO, and Bi-LSTM methods achieved averages of 87.46%, 73.22%, and 70.03%, respectively, all below the proposed method. In conclusion, the dual-sensor LM digital media game interaction system training method is more effective at capturing human motion information during interactions and providing timely feedback. This study did not include a detailed analysis of the game interaction system's language text parsing ability.

Therefore, future research can focus on further exploring human-computer interaction dialogues.

Funding

The research is supported by the General Project of the 2025 Annual Program under the 14th Five-Year Plan for Educational Science Research in Shaanxi Province, Research on the Teaching Reform Path of Private Universities in Shaanxi Based on the "Interest-Knowledge" Dual Mapping, Project Number: SGH25Y3254.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The author confirms that no AI tools were used in the preparation of this manuscript.

Reference

Avcı, Ü., and Akgül, F. (2024). The effect of metaverse technologies on education and human–computer interaction: A systematic analysis of the literature. *Journal of Educational Technology and Online Learning,* 7(1), 1–16.

Fischer, A., Izmailov, A. F., and Solodov, M. V. (2024). The Levenberg–Marquardt method: An overview of modern convergence theories and more. *Computational Optimization and Applications,* 89(1), 33–67.

Gil, R. M., Gutiérrez-Ujaque, D., and Teixidó, M. (2024). Analyzing the metaverse: Computer games, blockchain, and 21st-century challenges. *International Journal of Human–Computer Interaction,* 40(21), 6758–6775.

Guo, Z., and Lo, C. H. (2024). An empirical framework for understanding a player's sense of agency in games. *International Journal of Human-Computer Interaction,* 40(19), 5717-5736.

Hakim, G. P. N., Hadi Habaebi, M., Elsheikh, E. A. A., Suliman, F. M., Islam, M. R., Yusoff, S. H. B., and Anzum, R. (2024). Levenberg–Marquardt artificial neural network model for self-organising networks implementation in wireless sensor networks. *IET Wireless Sensor Systems,* 14(5), 195–208.

Hare, R., Ferguson, S., and Tang, Y. (2025). Enhancing student experience and learning with iterative design in an intelligent educational game. *British Journal of Educational Technology,* 56(2), 551–568.

Liao, S. H., Widowati, R., and Tang, W. (2025). Social media, mobile payment, and mobile gaming for intentional and behavioral recommendations. *International Journal of Human–Computer Interaction,* 41(4), 2560–2578.

Lyu, Z. (2024). State-of-the-art human–computer interaction in metaverse. *International Journal of Human–Computer Interaction,* 40(21), 6690–6708.

Onyejelem, T. E., and Aondover, E. M. (2024). Digital generative multimedia tool theory (DGMTT): *A theoretical postulation. Journalism,* 14(3), 189–204.

Revanesh, M., Gundal, S. S., Arunkumar, J. R., Josephson, P. J., Suhasini, S., and Devi, T. K. (2024). Artificial neural networks-based improved Levenberg–Marquardt neural networks for energy efficiency and anomaly detection in WSN. *Wireless Networks,* 30(6), 5613–5628.

Saputri, D. Y., and Indriayu, M. (2024). Learning transformation 4.0 with games-based interactive multimedia: Its effectiveness towards elementary school students' cognitive skills. *Pegem Journal of Education and Instruction,* 14(4), 381–392.

Srilakshmi, K., Rao, G. S., Swarnasri, K., Inkollu, S. R., Kondreddi, K., Balachandran, P. K., and Colak, I. (2024). Optimization of ANFIS controller for solar/battery sources fed UPQC using a hybrid algorithm. *Electrical Engineering,* 106(4), 3743–3770.

Sun, L., and Kwak, N. (2024). Multimedia human–computer interaction method in video animation based on artificial intelligence technology. *International Journal of Information Technology and Web Engineering,* 19(1), 1–15.

Uwimana, E., Zhou, Y., and Sall, N. M. (2025). A short-term load demand forecasting: Levenberg-Marquardt (LM), Bayesian regularization (BR), and scaled conjugate gradient (SCG) optimization algorithm analysis. *The Journal of Supercomputing,* 81(1), 55.1-55.30.

Vilar, J., Rodrigues, A., and Correia, N. (2025). An extended reality platform for cultural gaming: Enabling interactive narratives in spatial contexts. *ACM Journal of Computing and Cultural Heritage*, 17(4), 1–18.

Yang, Y., and Bao, W. (2025). Application of human–computer interaction technology in remote language learning platform. *International Journal of Human–Computer Interaction,* 41(3), 1751–1761.

Yokoi, R., and Nakayachi, K. (2025). Human motivation in competition against artificial intelligence: Using one-to-one games. *International Journal of Human–Computer Interaction,* 41(4), 2198–2210.

Yuzer, E. O., and Bozkurt, A. (2024). Instant solar irradiation forecasting for solar power plants using different ANN algorithms and network models. *Electrical Engineering,* 106(4), 3671–3689

Zargham, N., Friehs, M. A., Tonini, L., Alexandrovsky, D., Ruthven, E. G., Nacke, L. E., and Malaka, R. (2025). Let's talk games: An expert exploration of speech interaction with NPCs. *International Journal of Human–Computer Interaction,* 41(5), 3592–3612.

Zhang, J., Tian, J., Alcaide, A. M., Leon, J. I., Vazquez, S., Franquelo, L. G., and Yin, S. (2023). Lifetime extension approach based on the Levenberg–Marquardt neural network and power routing of DC–DC converters. *IEEE Transactions on Power Electronics,* 38(8), 10280–10291.

Yu Ma received her master's degree in art studies from Guangxi Normal University in 2014. Currently, she is an Associate Professor and full-time teacher in the Department of Digital Media Technology at Xi'an Technology and Business University, and additionally serves as a college-level Supervisor within the university's School of Information Engineering. Her work focuses extensively on digital art teaching and research, covering core courses such as Web Design and Production, 2D Digital Animation Design, and Creative Design. Her mentorship has led students to achievements, including the Provincial Third Prize in the National College Student Advertising Art Competition and the Third Prize in the National College Student Digital Media Technology Works and Creative Upgrade Competition, while she herself has received recognition as a University-level Excellent Teacher and an Excellent Instructor in the National College Student Advertising Art Competition. In research, she has presided over three provincial and university key projects and led two horizontal projects. She has published eight academic papers, obtained 2 utility model patents, and her relevant research achievements have won the Second Prize in Shaanxi Provincial Art Education Scientific Research Papers. Her research interests include integrating immersive art education and digital technology, the digital artistic expression of traditional cultural elements, and the application of digital media technology.