International Journal of Mechatronics and Applied Mechanics
2025, Issue 21, Vol. 1

ENHANCING VIRTUAL REALITY PRODUCT INTERACTION THROUGH REAL-TIME 3D SCENE PERCEPTION AND MULTI-MODAL INTERACTION

Qianqian Hu[1], Heguang Wang[2*] and Bing Wang[3]
[1]School of Art and Design, Zhengzhou University of Light Industry, Zhengzhou, 450000, China
[2]School of Cultural Heritage and Art Design, Zhengzhou University of Technology, Zhengzhou, 450000, China
[3]Rehabilitation Engineering College, China Civil Affairs University, Beijing, 102600, China

Abstract - The advancement of technology has spurred the development of numerous techniques for simulating real products in 3D virtual environments. However, traditional Virtual Reality (VR) product interaction systems mainly promote interaction in 2D scenes, resulting in unsatisfactory interaction experiences in 3D settings. To address this issue, this study proposes a new design method for a VR product interaction system that utilizes real-time 3D scene perception. This method constructs a 3D model through 3D scene perception and extracts relevant object information. Multi-modal interaction techniques are employed to gather product interaction data from users' various sensory inputs. The experimental results showed that the proposed system achieved a low system occupancy rate of 0.06%, which was superior to other VR product interaction systems. With the increase of product interaction data, the packet loss rate remained stable at 0.053%, ensuring the smooth operation of the system. These research results indicate that VR product interaction systems based on 3D scene perception can effectively achieve interaction with 3D scene products, thereby improving user satisfaction and immersion.

Keywords: 3D scene; Multi-modal interaction; Virtual reality; Product interaction system; Real-time perception.

1. Introduction

The continuous improvement of technological level has promoted the maturity of sensor technology and digital analysis technology. The rapid development of these technologies has driven the continuous advancement of Virtual Reality (VR) technology [1]. Under the influence of today's information age, the product and user experience are becoming increasingly complex, and traditional product interaction methods cannot meet the needs of users well. Therefore, VR product interaction has emerged [2]. VR technology greatly improves user satisfaction with product interaction by transforming real environments into virtual scenes [3]. With the increasingly widespread application of VR product interaction in daily life, designing a VR Product Interaction System (VR-PIS) has become a huge challenge at present. Traditional VR-PIS designs often use cameras to construct virtual scenes. Users interact with virtual products through computers. These traditional VR-PIS can improve the convenience of user interaction with products [4]. However, with the continuous increase of virtual scene data, these traditional system design methods cannot effectively design VR-PIS and require a large amount of software and hardware for the system.

Therefore, there is an urgent need for a method to design VR-PIS. When interacting with VR products, users need to interact with virtual products through different senses. The use of Multi-modal Interaction (MI) enables users to interact with computers through different modes, providing users with a more natural and efficient interactive experience [5]. However, using MI for VR-PIS design cannot produce products in 3D form, resulting in poor user experience and realism. Real-time 3D Scene (RT3DS) perception is capable of transforming 2D scenes into 3D scenes and reproducing virtual environments in a realistic way [6]. The perception of RT3DS greatly improves the realism of user experience, making up for the shortcomings of MI in VR-PIS design.

Therefore, this study introduces RT3DS perception combined with MI for the design of VR-PIS, aiming to improve user satisfaction with VR-PIS.

dx.doi.org/10.17683/ijomam/issue21.11

2. Related Work

With the gradual increase in the application scope of VR technology, VR product interaction has received widespread attention from industry scholars. Liu X et al. proposed a VR method based on sensory engineering to address the issue that sample images and line sketches cannot fully express the true form of products. It constructed the mapping relationship between products and users from four dimensions: overall, unit, interrelationships, and details, and used multiple linear regression analysis to analyze the correlation between products and perceived images [7]. Zhao L et al. proposed a VR-based rehabilitation service system design method to address inaccurate evaluation of rehabilitation outcomes in most rehabilitation service systems.

This method established a virtual training scenario optimization mechanism for rehabilitation effectiveness, as well as a knowledge graph of rehabilitation plans, and established a service system framework that meets user needs [8]. Ahn D K et al. proposed an interactive VR fashion show design method to address low user experience in video-based digital fashion show stages. It selected elements that affect the interactive user experience and constructed a system framework based on the scope of interactive user experience applications, using commercial game engines for process simulation [9]. Suhaimi N S et al. proposed an emotion recognition method based on VR combined with electroencephalography to identify user emotions. This method constructed a VR database, used VR data to stimulate user emotions, conducted EEG for user immersion experience, and evaluated the emotion recognition system through machine learning algorithms [10]. Melo M et al. proposed an immersive VR tourism based on VR technology to provide a multi-sensory experience for tourism. This plan involved constructing a tourism area model and utilizing sensors to collect virtual sensory information data. The proposed method achieved a satisfaction rate of 90.12% for user experience [11].

In addition, research on 3D scene perception has also received widespread attention from many scholars. Maccarone et al. proposed an RT3DS reconstruction method based on single photon detection for radar imaging of underwater scattering. It constructed detector arrays through complementary metal semiconductor technology and measured photon flight time using resolution time. This method achieved an image processing time of 33ms [12]. Hughes N et al. proposed a 3D scene construction perception system based on real-time spatial perception to address the difficulty of constructing rich 3D scene maps. The system gradually constructed scene layers through real-time algorithms, established local Euclidean distance functions, extracted topology maps from the distance functions, and used real-time algorithms to optimize 3D scene maps for loop closure [13]. Yang B et al. proposed a 3D scene rendering method based on closed neural rendering to address the difficulty of understanding scenes from any angle using computers. This method learned scene object data through offline learning, estimated the correct 3D scene layout using overall neural rendering, and constructed an optimization framework for neural rendering [14]. Kong L et al. proposed a scene construction method based on multi-modal data to address the high dependence of 3D scenes on LiDAR in autonomous driving. It constructed a framework for efficient data learning using multi-modal data by integrating laser beam operations scanned by different LiDAR and combining the corresponding relationship between LiDAR and camera [15]. Ding R et al. proposed a language-driven 3D scene understanding method to address insufficient understanding of open world instance-level scenes in traditional 3D scenes. It used a trained visual language model for image text encoding and generated subtitles for multi-view images of 3D scenes, and designed a hierarchical point association method to learn semantic perception embedding [16].

In dynamic gesture recognition, the original depth image has problems such as low contrast in the area of interest of the hand, difficulty in highlighting details such as finger direction and finger palm overlap. Mahmud et al. proposed using quantized depth images instead of original depth images as input modalities, which improved the performance of gesture recognition [17]. Traditional wearable human motion recognition methods only focus on user motion recognition, lacking interactivity and user engagement. Xiao proposed a new immersive HAR method called MovPos VR, which utilizes VR technology to create realistic scenes to enhance user experience [18].

In summary, many traditional methods rely on static scenes and manual control interactions, and fail to fully utilize RT3DS perception and MI, resulting in limited effectiveness under dynamic user demands. Secondly, although some studies have made progress in 3D scene reconstruction, real-time performance and computational burden are still bottlenecks, limiting their practical applications. In addition, most existing experiments are limited to small-scale simulations and fail to cover various practical application scenarios, ignoring the differences in system adaptability among different user groups. Relatively speaking, there are still many gaps in the systematic exploration of the MI combination and 3D scene perception. Therefore, this study introduces RT3DS perception combined with MI for VR-PIS design, aiming to improve users' experience of VR product interaction.

3. VR-PIS Design Integrating 3D Scenes

3.1 Construction of VR-PIS based on MI

VR interaction refers to a virtual environment that does not exist in reality through computer technology. Different users interact with the established virtual environment through their visual, auditory, and tactile senses, creating a sense of immersion for them [19]. The interaction process of users in the virtual scene is shown in Fig. 1.

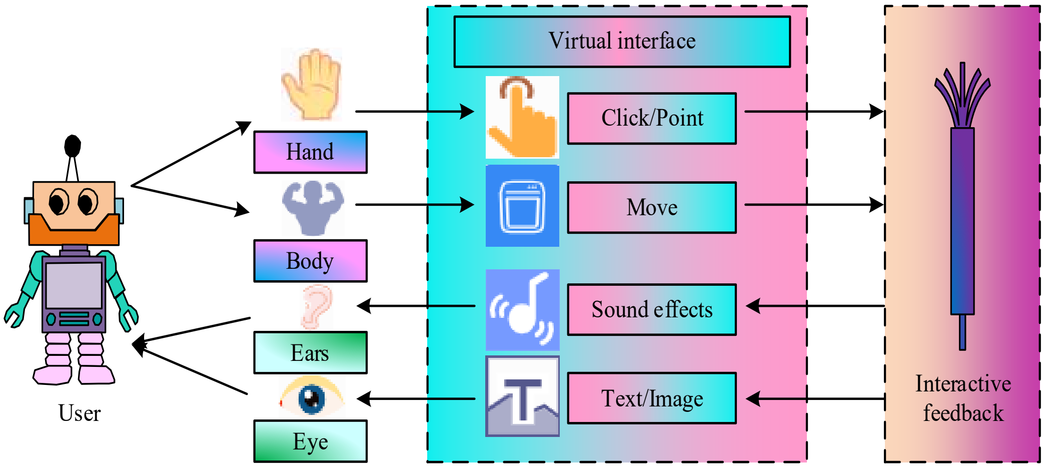


Figure 1: Virtual scene interaction process

In Fig. 1, the user clicks and moves within the virtual interface using their hands and body. The virtual interface provides feedback through user actions, delivering sound effects and text images to the user's ears and eyes respectively. Users reproduce their real-life experiences in virtual scenes through their senses such as ears and eyes. In traditional virtual scenes, interaction is only conducted through a single sensory input. MI refers to the process of interaction between humans and computers through various input and output modes, including voice, image, text, and other modes. This interaction method aims to enhance the user experience, enabling computers to understand and process different types of input information, thereby providing a more natural and efficient interaction experience. To collect the position of objects in the VR scene, a Multimodal Fusion (MF) algorithm is used to calculate the object information. By using weighted methods, visual, auditory, and tactile inputs are integrated to provide a comprehensive representation of user interaction. The mathematical expression of MF is shown in equation (1).

$$do_i = \sqrt{(x_g - x_i)^2 + (y_g - y_i)^2 + (z_g - z_i)^2} \tag{1}$$

In equation (1), do_i is the distance between the virtual character and the object. x_i, y_i, and z_i are the axis coordinate points of the object in VR space. When calculating the confidence level of character actions, it can be calculated through the information entropy equation, as shown in equation (2).

$$H(G_i) = 1 + \sum_{i=1}^{n} P(G_i) \log P(G_i) \tag{2}$$

In equation (2), $H(G_i)$ is the confidence level, and G_i is the recognition probability of the object. P representing probability. n represents the number of elements involved in the calculation. When performing speech recognition, the mathematical formula is shown in equation (3).

$$H(V_i) = 1 + \sum_{i=1}^{n} P(V_i) \log P(V_i) \tag{3}$$

In equation (3), $H(V_i)$ is the fitness value for speech recognition. V_i is the probability of matching received speech. In VR, MI mainly analyzes the perception of information through the senses of characters. The dimensions of human sensory and modal recognition are shown in Fig. 2.

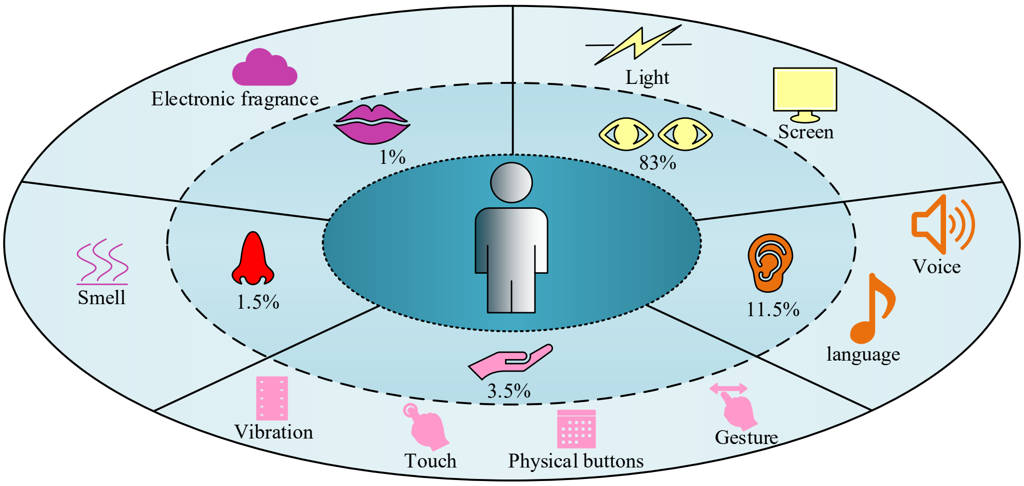


Figure 2: Sensory and modal dimensions

In Fig. 2, when the human body undergoes MI, it mainly interacts through vision. Visual perception interacts with the perception of the screen and light. Secondly, modal interaction is carried out through auditory perception, where external sounds and language information are received through the ears for interaction. During the interaction process, the frequency of using senses such as the mouth and nose is relatively low, while using hands for interaction is relatively more frequent compared to using senses such as the mouth and nose. The distance probability calculation for the position of objects in VR space is shown in equation (4).

$$P(D_i) = \frac{do_i}{\sum_{i=1}^{n} do_i} \tag{4}$$

In equation (4), $P(D_i)$ is the probability of the object's appearance position distance in VR space. The expression of weighting calculation using equations based on confidence values of different modalities is shown in equation (5).

$$\begin{cases} \partial_G = \dfrac{H(G_i)}{H(G_i)+H(V_i)+H(D_i)} \\ \partial_V = \dfrac{H(V_i)}{H(G_i)+H(V_i)+H(D_i)} \\ \partial_D = \dfrac{H(D_i)}{H(G_i)+H(V_i)+H(D_i)} \end{cases} \tag{5}$$

In equation (5), ∂_G, ∂_V, and ∂_D are weight values for character actions, speech, and distance. When interacting with products in VR space, users will approach a certain item. When the user's desired product does not appear in the VR space visible to the user, the mathematical form of the distance between the product and the user is shown in equation (6).

$$A_{Q_i} = \frac{\sum_{i=1}^{n} do_i}{n} \tag{6}$$

In equation (6), A_{Q_i} is the relative distance between the product and the user. The structure of VR-PIS using multi-module interaction is shown in Fig. 3.

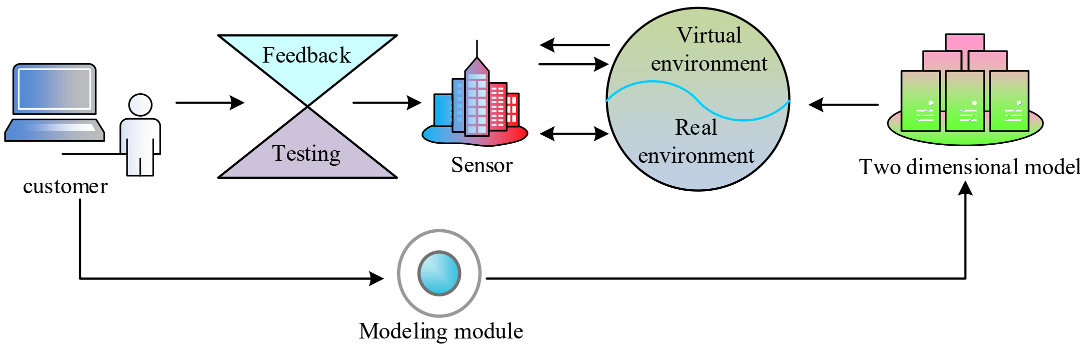


Figure 3: Construction of interactive system for VR products

In Fig. 3, MI is used to construct VR-PIS. Sensors are used to transmit information when virtual users enter the system. The virtual environment integrates information through the MI module when receiving transmission signals [20]. The information is integrated and transmitted to multi-modal sensors, and ultimately fed back to the user.

3.2 Construction of VR-PIS Integrating 3D Scene Perception

The VR system constructed using MI can effectively interact with different products, and the product interaction effect in 2D space is good. However, the constructed VR-PIS cannot effectively interact with VR products in 3D scenes. Therefore, it has been studied to introduce 3D scene perception for the construction of VR-PIS. 3D scene perception can transform 2D planes into 3D and restore virtual scenes to reality [21-22]. The construction process of the 3D scene is shown in Fig. 4.

In Fig. 4, when constructing a 3D scene, the first step is to collect environmental data information, collect environmental satellite images, and convert the collected data information into a terrain database through texture mapping. The next step is to collect structural information of objects in the environment, organize the data structure, and also perform texture mapping to create a ground object database. Finally, the 3D scene is constructed based on the created terrain and terrain database. In the constructed 3D scene, the camera spatial coordinates can be expressed mathematically, as shown in equation (7).

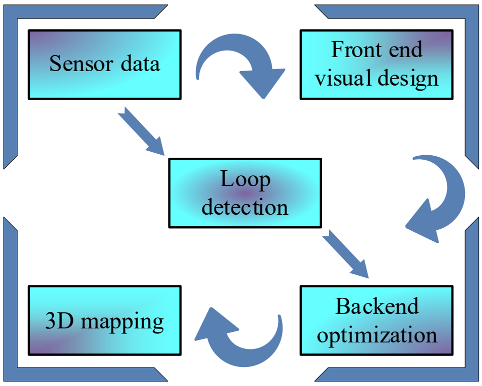


Figure 4: 3D scene construction process

$$\begin{cases} x_i = \dfrac{1}{z_c} \cdot f \cdot x_c \\ y_i = \dfrac{1}{z_c} \cdot f \cdot y_c \end{cases} \tag{7}$$

In equation (7), x_i and y_i are spatial coordinate points, and f is the focal length. The positional relationship between the 3D scene camera and spatial coordinates is shown in equation (8).

$$s_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = KT \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{8}$$

In equation (8), K is the parameter of the camera, and T is the transformation matrix in the 3D image. The projection error problem in 3D scenes can be represented by a Gaussian distribution in 3D space, as shown in equation (9).

$$T^* = \min_x \sum \|e_i\|^2 \tag{9}$$

In equation (9), e_i is the projection error from the 3D spatial coordinate point to the plane mapped by the camera.

The construction process of the PIS by combining RT3DS perception and VR is shown in Fig. 5.

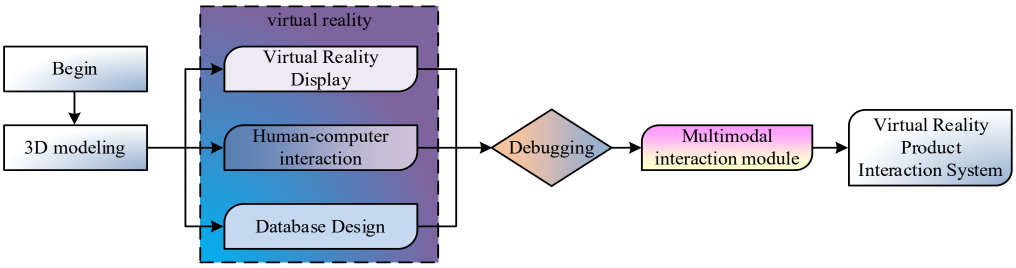


Figure 5: Construction process of VR-PIS

In Fig. 5, the construction of the interactive system is achieved through RT3DS for the realistic reproduction of the real environment, collecting object information from the real environment and transmitting it to the virtual environment. Environmental data such as depth information and texture information in the scene are collected through a variety of sensors (LiDAR, RGB camera).

After preprocessing, the collected data are used to generate a 3D scene model through the real-time 3D reconstruction algorithm voxel grid method. On this basis, real-time display of 3D scenes through optimized rendering engines ensures high frame rates, low latency, and smooth user experience. To extract geometric feature information of objects in 3D scenes, 3D scenes are referred to as model objects. The geometric information feature measurement of objects in the model is shown in equation (10).

$$E_g^k(v) = E_n^k(v) + E_f^k(v) \qquad (10)$$

In equation (10), $E_n^k(v)$ is the proportion of objects in the 3D scene model to the center point in 3D space, and $E_f^k(v)$ is the feature distribution of the measured object. The metric function of a 3D scene model can be expressed by geometric information and feature information, as shown in equation (11).

$$E_m(v) = \sum_{k=1}^{K} w_k E_g^k(v) = \sum_{k=1}^{K} w_k (E_n^k(v) + E_f^k(v)) \qquad (11)$$

In equation (11), w_k is the structural feature of the 3D scene model. The basic framework for constructing a system that integrates RT3DS perception for VR product interaction is shown in Fig. 6.

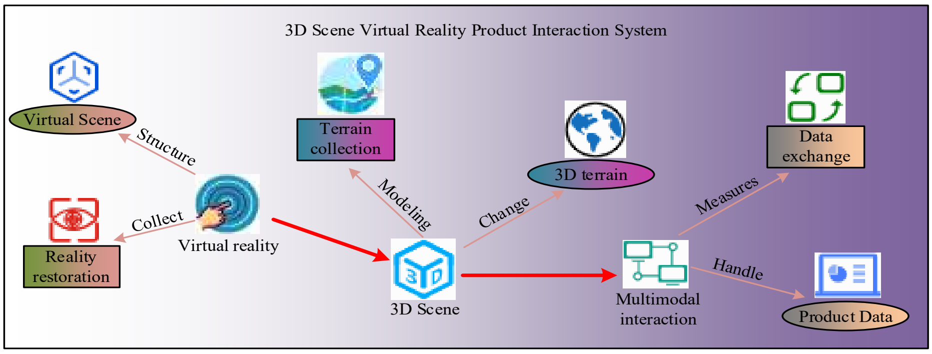


Figure 6: Basic framework of VR-PIS integrating 3D scenes

In Fig. 6, the structure is mainly composed of 3D scenes, sensors, and product interaction modules. The perception of RT3DS mainly involves converting 2D images to 3D and constructing a 3D model framework. Sensors interact with product data and transmit user actions in VR to the interactive system. Users can choose products based on their own needs in the VR-PIS perceived by RT3DS. To obtain user demand weights, the system can calculate them through quality function deployment theory, as shown in (12).

$$DR_j = \sum_{i=1}^{n} CR_i R_{ij} \qquad (12)$$

In equation (12), CR_i is the weight of the demand for items required by different users. R_{ij} is the assignment between the weight of objects required by different users and the weight of product design. DR_j is the design weight of the product required by the user.

4. Empirical Analysis of 3D Scene Perception Understanding VR-PIS

4.1 Performance analysis based on 3D and VR-PIS

To analyze the performance of RT3DS in perceiving VR-PIS, this study conducts simulation experiments in MATLAB R2024b and ExtendSim 2024. Among them, ExtendSim is used to simulate the data flow of product interaction and conduct system performance evaluation. Table 1 shows the experimental configuration parameters.

Table. 1 Experimental Environment

Project	Parameter model
System	Windows 11
CPU	Intel i5-11300H
Graphics card	AMD RX 7600
Simulation software	ExtendSim
Software version	ExtendSim 2024

The different VR-PIS design methods collected from the internet include those based on Developmental Movement Studies (DMS) and those based on Visual Communication (VC). A total of 200 sets of data are collected from different products in the network, of which 150 sets are used as test data and the other 50 sets are utilized as training data to construct the experimental dataset. To validate the performance of the proposed method's system Packet Loss Rate (PLR), it is compared with DMS and VC methods, as shown in Fig. 7.

In Fig. 7 (a), under the condition of 50 datasets, the average PLR of the research system is only 0.054%, significantly lower than other methods, and the variation amplitude of the system PLR is relatively gentle. In DMS and VC, the lowest system PLR is 0.125%. In Fig. 7 (b), among the 100 datasets, the average PLR of the research method is only 0.053%, while the average PLR of the DMS system reaches 0.197%. The average PLR of VC is slightly lower than DMS, but significantly higher than the research method. In Fig. 7 (c), under 150 conditions, the performance of the research system is optimal, with an average PLR of 0.051%. The research method has a lower system PLR and has verified the good performance advantages of the system. Afterwards, to verify the System Occupancy Rate (SOR) of the designed system, SOR comparisons are conducted under the same dataset conditions, as shown in Fig. 8.

In Fig. 8 (a), among 50 datasets, the three methods show significant advantages in comparing SOR, with an average SOR of 0.06% and no significant changes in their average occupancy rate. The lowest average SOR of the other two methods is 0.10%, significantly higher than the highest SOR of the research method. In Fig. 8 (b), among the 100 experimental datasets, the average SOR of the research system is 0.12%, while the average occupancy rates of DMS and VC are significantly higher than those of the research system. In Fig. 8 (c), the average SOR of the research method is 0.15% in 150 datasets. The average market share of VC is 0.27%, slightly lower than DMS, but significantly higher than the research method. Therefore, the research system has significant advantages over other methods and has the lowest market share. To further validate the performance of 3D scene aware VR-PIS, a comparison of system response time is conducted on the research system, as shown in Fig. 9.

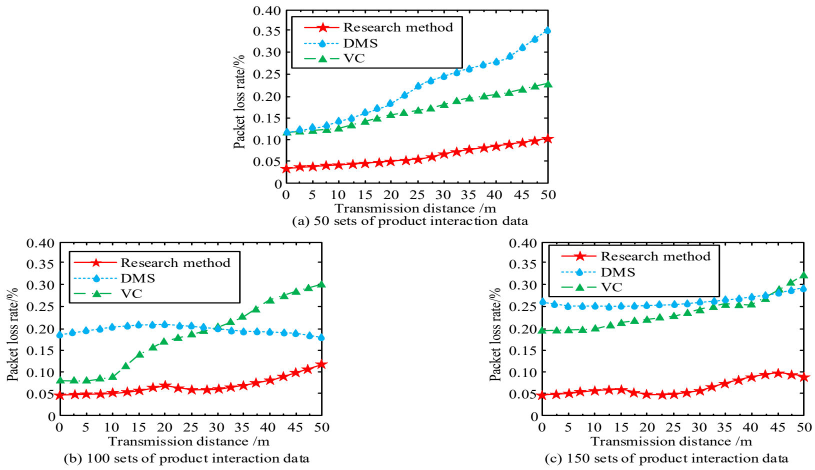


Figure 7: Comparison of PLRs in systems with different design methods

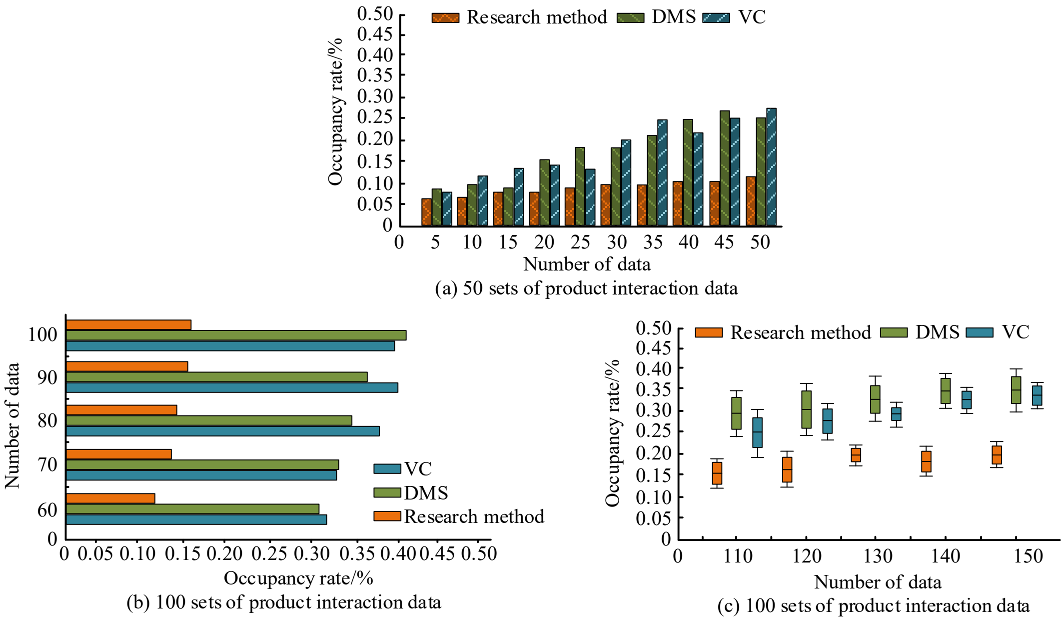


Figure 8: Comparison of SOR using different designed methods

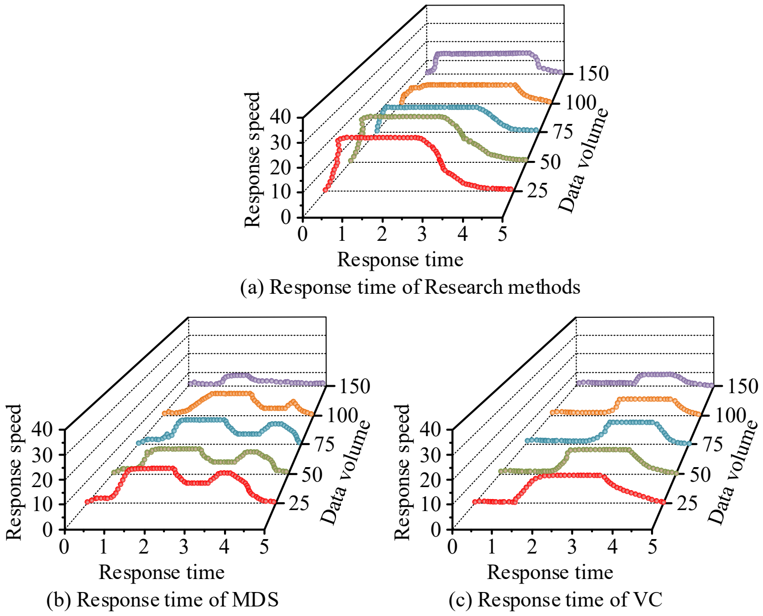


Figure 9: Comparison of response times of different systems

In Fig. 9 (a), the response speed of the research system is the fastest under the condition of 25 experimental datasets, and slightly slower under the condition of 100 datasets. The average response speed of the system shows a slow downward trend with increasing response time. This indicates that the research system has a faster response speed in VR products. In Fig. 9 (b), the response speed of the DMS system is slow and shows a slow downward trend with increasing response time. The response speed of DMS is relatively fast, but under the condition of 25 datasets, its fastest response speed is far inferior to that of the research system. In Fig. 9 (c), the response speed of the VC method is the slowest compared to the other two systems. In summary, the three systems showed significant advantages when facing VR products, and their average response speed was the fastest.

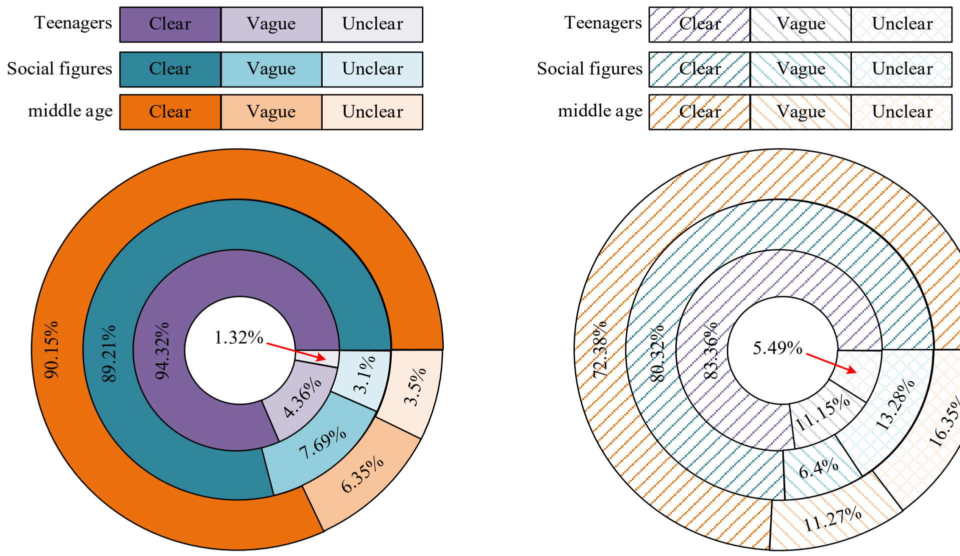
4.2 Practical Effect Analysis of 3DVR-PIS

To verify the practical application effect of RT3DS perception of VR-PIS, this study collects data from different products in the network, including gaming, television, and shopping data. Users use 3D scene perception VR interactive systems to interact with different products. Table 2 shows the satisfaction results of the research system.

In Table 2, among the three populations of adolescents, social workers, and students, the satisfaction rate of the research system with different VR-PIS is the highest, with an average satisfaction rate of over 91.35%. Its satisfaction is 6.32% and 8.56% higher than other systems, respectively. The comparison of the flexibility and convenience of different operating systems when users use the 3D scene perception VR interactive system is shown in Fig. 10.

Table. 2 Comparison of user satisfaction with different interactive systems

User	Product	Research system satisfaction	DMS system satisfaction	VC system satisfaction
	Game	94.24%	90.11%	84.63%
Teenagers	Television	92.31%	87.64%	84.21%
	Shopping	90.03%	89.26%	83.19%
	Game	93.65%	89.57%	86.57%
Social figures	Television	89.05%	84.36%	84.47%
	Shopping	91.27%	85.49%	83.11%
	Game	88.63%	79.69%	72.59%
Student	Television	94.32%	84.67%	81.28%
	Shopping	94.78%	85.36%	80.36%



(a) Adaptability of different populations to this research method

(b) Adaptability of different populations to MDS interactive system

Fig. 10 Comparison of user flexibility in operating different systems

In Fig. 10 (a), different users have a higher overall proficiency in using the research system to perceive VR in 3D scenes. Among them, teenagers have a relatively flexible use of the system, and the proficiency rate has reached 94.32%. Only a small number of people are unable to flexibly apply the system. Middle-aged individuals and members of society have higher flexibility in using research systems, with average proficiency rates of 90.15% and 89.21%, respectively. This indicates that different users have higher user satisfaction and flexibility when using the system to simulate 3D scenes in reality. In Fig. 10 (b), different users have lower overall flexibility when using the MDS construction system to perceive 3D scenes in VR. Among them, middle-aged people cannot clearly apply the system flexibly when using it to simulate reality. In summary, different users can use the research system more clearly and flexibly to simulate reality in 3D scenes when using different systems.

4.3 Real Case Study: Game Application

The study collaborates with a VR game company and selects 10 players to participate in the experiment, aiming to evaluate the user experience of the VR-PIS system in VR games. The experimental design requires players to interact with objects in the virtual game environment through gesture recognition and voice commands. Tasks include object grabbing, combat actions, etc. To comprehensively evaluate the system performance, data collection is carried out through the following indicators: system response time, frame rate, input delay, action accuracy, number of virtual object interactions, game task completion, and player physiological feedback. The specific results are shown in Table 3.

Table. 3 Game application scenario - System performance and player feedback

Player number	System response time (s)	Frame rate (FPS)	Input delay (s)	Action accuracy (%)	Virtual object interaction count	Task completion (%)	Average heart rate (beats/min)	Skin conductance (μS)
1	0.38	62	0.25	92.45	42	85.55	78	5.44
2	0.34	59	0.3	87.13	39	79.23	82	4.89
3	0.29	60	0.28	90.32	45	88.33	80	5.12
4	0.41	55	0.35	85.78	41	76.47	75	6.03
5	0.36	58	0.32	88.65	43	82.13	77	5.34
6	0.33	61	0.27	91.22	47	90.52	79	5.01
7	0.37	57	0.33	86.47	38	78.82	81	4.72
8	0.4	63	0.31	88.72	40	83.33	76	5.21
9	0.35	60	0.29	90.84	44	87.05	80	5.42
10	0.39	56	0.36	84.96	36	75.78	83	5.63

As shown in Table 3, the system response time fluctuates between 0.29 s and 0.41 s, and the frame rate is 55-63 FPS, indicating that the system is stable under different player operations. The combination of input delay and action accuracy can further illustrate the effectiveness of the proposed system. The number of interactions between players and virtual objects is 36-47 times, and the average task completion rate is 83.18%, reflecting the effectiveness of the system in promoting user participation and task execution. The average heart rate and skin galvanic response of players show that VR-PIS is stable in terms of interactivity, immersion, and user participation in the game.

5. Conclusions

The continuous improvement of VR technology has made product interaction in virtual scenes increasingly common. VR-PIS for 3D scenes has become a huge challenge at present. This study proposed a VR-PIS based on RT3DS perception to address the issue of poor interaction effects of traditional VR interaction systems in 3D scenes. A 3D scene model was constructed using RT3DS perception, and MI was utilized to build the product interaction module. In the experiment, the proposed VR-PIS showed an average occupancy rate of 0.12% for the system under multiple interactive data conditions, significantly lower than the interactive systems based on MDS and VC, and the response time of the system was shorter. In the practical application of the research system, teenagers, students, and members of society had the highest average satisfaction with the research system, reaching 91.35%, which was 6.32% and 8.56% higher than other systems. Moreover, according to the proficiency level of different groups in using the interactive system, the research system's proficiency level for teenagers reached 94.32%. The proficiency levels of middle-aged individuals and members of society reached 90.15% and 89.21%, respectively. In summary, the VR-PIS based on RT3DS can improve the user experience for different users.

Although the VR-PIS system performs well in multiple application scenarios, it still has some limitations. First, the system has a high demand for computing resources, which may affect performance in complex virtual scenes. Second, although gesture recognition and voice interaction have made progress, the accuracy in dynamic environments still needs to be improved. In addition, the current experiments are mainly focused on ordinary user groups, and the sample range needs to be expanded in the future.

Future research will focus on optimizing algorithms to reduce computing resource consumption and improve the system's operating efficiency on low-configuration devices. In addition, artificial intelligence technology is combined to improve interaction accuracy, especially recognition capabilities in dynamic environments. At the same time, the system will be applied to a diverse user group to ensure its wide adaptability and further improve the practicality and scalability of the system.

References

[1] Enyejo J O, Obani O Q, Afolabi O, Igba E, Ibokette A I. Effect of Augmented Reality (AR) and Virtual Reality (VR) experiences on customer engagement and purchase behavior in retail stores. Magna Scientia Advanced Research and Reviews, 2024, 11(2): 132-150.

[2] Luangrath A W, Peck J, Hedgcock W, Xu Y. Observing product touch: The vicarious haptic effect in digital marketing and virtual reality. Journal of Marketing Research, 2022, 59(2): 306-326.

[3] Harz N, Hohenberg S, Homburg C. Virtual reality in new product development: Insights from prelaunch sales forecasting for durables. Journal of Marketing, 2022, 86(3): 157-179.

[4] Rauschnabel P A, Felix R, Hinsch C, Shahab H, Alt F. What is XR? Towards a framework for augmented and virtual reality[J]. Computers in human behavior, 2022, 133(6): 276-291.

[5] Lindell A E, Zimmermann-Kogadeeva M, Patil K R. Multimodal interactions of drugs, natural compounds and pollutants with the gut microbiota. Nature Reviews Microbiology, 2022, 20(7): 431-443.

[6] Zhang X, Bai X, Zhang S, He W, Wang P, Wang Z, Yu Q. Real-time 3D video-based MR remote collaboration using gesture cues and virtual replicas. The International Journal of Advanced Manufacturing Technology, 2022, 121(11): 7697-7719.

[7] Liu X, Yang S. Study on product form design via Kansei engineering and virtual reality. Journal of Engineering Design, 2022, 33(6): 412-440.

[8] Zhao L, Zhao Y, Bu L,Sun H, Tang W, Li K, Zhang Y. Design method of a smart rehabilitation product service system based on virtual scenarios: A case study. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31(3): 4570-4579.

[9] Ahn D K, Bae B C, Kim Y. User Experience of a Digital Fashion Show: Exploring the Effectiveness of Interactivity in Virtual Reality. Applied Sciences, 2023, 13(4): 2558-2571.

[10] Suhaimi N S, Mountstephens J, Teo J. A dataset for emotion recognition using virtual reality and EEG (DER-VREEG): Emotional state classification using low-cost wearable VR-EEG headsets. Big Data and Cognitive Computing, 2022, 6(1): 16-37.

[11] Melo M, Coelho H, Gonçalves G, Losada N, Jorge F, Teixeira M S, Bessa M. Immersive multisensory virtual reality technologies for virtual tourism: A study of the user’s sense of presence, satisfaction, emotions, and attitudes. Multimedia Systems, 2022, 28(3): 1027-1037.

[12] Maccarone A, Drummond K, McCarthy A, Steinlehner U K, Tachella J, Garcia D A, Buller G S. Submerged single-photon LiDAR imaging sensor used for real-time 3D scene reconstruction in scattering underwater environments. Optics Express, 2023, 31(10): 16690-16708.

[13] Hughes N, Chang Y, Carlone L. Hydra: A Real-time Spatial Perception System for 3D Scene Graph Construction and Optimization. arXiv e-prints, 2022, 22(1): 13360-113369.

[14] Yang B, Zhang Y, Li Y, Cui Z, Fanello S, Bao H, Zhang G. Neural rendering in a room: amodal 3d understanding and free-viewpoint rendering for the closed scene composed of pre-captured objects. ACM Transactions on Graphics (TOG), 2022, 41(4): 1-10.

[15] Kong L, Xu X, Ren J, Zhang W, Pan L, Chen K, Liu Z. Multi-modal data-efficient 3d scene understanding for autonomous driving. arXiv preprint. 2024, 24(5): 52-69.

[16] Ding R, Yang J, Xue C, Zhang W, Bai S, Qi X. Lowis3d: Language-driven open-world instance-level 3d scene understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 8517-8533.

[17] Mahmud H, Morshed MM, Hasan MK. Quantized depth image and skeleton-based multimodal dynamic hand gesture recognition. The Visual Computer. 2024, 40(1): 11-25.

[18] Xiao Z, Chen Y, Zhou X, He M, Liu L, Yu F, Jiang M. Human action recognition in immersive virtual reality based on multi-scale spatio-temporal attention network. Computer Animation and Virtual Worlds. 2024 Sep;35(5): e2293.

[19] Villena-Taranilla R, Tirado-Olivares S, Cózar-Gutiérrez R, González-Calero J A. Effects of virtual reality on learning outcomes in K-6 education: A meta-analysis. Educational Research Review, 2022, 35(8): 434-445.

[20] Lian Z, Liu B, Tao J. Smin: Semi-supervised multi-modal interaction network for conversational emotion recognition. IEEE Transactions on Affective Computing, 2022, 14(3): 2415-2429.

[21] Yildirim I, Siegel M H, Soltani A A, Ray Chaudhuri S, Tenenbaum J B. Perception of 3D shape integrates intuitive physics and analysis-by-synthesis. Nature Human Behaviour, 2024, 8(2): 320-335.

[22] Zhang A, Eranki C, Zhang C, Park J H, Hong R, Kalyani P, Biswas J. Towards robust robot 3d perception in urban environments: The ut campus object dataset. IEEE Transactions on Robotics, 2024, 40(3): 3322-3340.