International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

To cite this article: Jiaxin Li, Jiaqi Lu, Qingzhuo Wang, Xiaomeng Jiang and Kehui Duo (2026). DESIGN OF MULTIMODAL INTERACTION-DRIVEN RAIL TRANSIT TRAINING SYSTEM: THEORETICAL FRAMEWORK AND APPLICATION RESEARCH, International Journal of Research in Commerce and Management Studies (IJRCMS) 8 (2): 477-486 Article No. 697 Sub Id 1119

DESIGN OF MULTIMODAL INTERACTION-DRIVEN RAIL TRANSIT TRAINING SYSTEM: THEORETICAL FRAMEWORK AND APPLICATION RESEARCH

Jiaxin Li, Jiaqi Lu, Qingzhuo Wang, Xiaomeng Jiang and Kehui Duo

Tianjin University of Technology, School of Management, Tianjin, 300384

DOI: https://doi.org/10.38193/IJRCMS.2026.8235

ABSTRACT

To address the problems of insufficient equipment, high operational risks, single interaction mode, and unbalanced distribution of educational resources in traditional training within the rail transit field, this study constructs a three-layer virtual simulation training system framework of “immersive scenarios-multimodal interaction-cloud data intelligence” based on embodied cognition and situational learning theories. The system framework restores rail transit equipment and scenarios at a 1:1 scale through three-dimensional modeling, integrates multimodal interaction methods including hardware sensing, mobile AR, voice/gesture recognition, and 5G remote holography, and realizes resource sharing and dynamic updates relying on cloud servers. It breaks through the temporal and spatial constraints of traditional training, not only reduces teaching costs, but also strengthens students' practical abilities and safety awareness through the operational experience in the training space. This framework effectively makes up for the deficiencies of existing teaching resources, provides an innovative solution for the cultivation of professional talents in the rail transit sector, and thus boasts significant teaching application value and promising promotion prospects.

KEYWORDS: Multimodal Interaction; Virtual Simulation Training System; Rail Transit

1. INTRODUCTION

Against the backdrop of the contradiction between the rapid development of digitalization and intelligence and the lag of traditional training methods, as well as the practical demand for urgently improving practitioners' safety awareness and emergency response capabilities, this paper puts forward the core problems existing in current rail transit training teaching, such as high equipment costs, single interaction mode, unbalanced teaching resources, and high practical operation risks. Traditional training relies on physical equipment, which suffers from pain points including "insufficient equipment for a large number of students", "invisibility of internal structures", and "low operational fault tolerance". Moreover, voice recognition has inadequate adaptability to professional terminology, and complex gesture recognition lacks real-time performance, making it difficult to meet

International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

diverse learning needs. To address these issues, this study deeply integrates cutting-edge technologies such as 5G communication, virtual reality (VR), augmented reality (AR), holographic technology, and the Internet of Things (IoT), and constructs a three-dimensional architecture of "immersive scenario layer - multimodal interaction layer - cloud data intelligence layer". By reproducing equipment and scenarios through 1:1 high-fidelity 3D modeling, developing multimodal interaction methods including hardware sensing, mobile terminals, voice/gesture recognition, and remote holography, and combining with cloud data storage and intelligent analysis, an innovative, efficient, and user-friendly virtual training environment is created. The research aims to break the temporal and spatial limitations and resource constraints of traditional teaching, provide an immersive, repeatable, and low-risk training experience, improve students' mastery of professional knowledge and practical operation capabilities, promote the balanced allocation of educational resources, and offer a brand-new solution for talent cultivation in the rail transit industry[i].

2. LITERATURE REVIEW

Domestic scholars have achieved fruitful results in the field of rail transit virtual simulation. Xie Rongli constructed a control model based on distributed VR technology to realize real-time simulation of train status and support for multiple operations[ii]. Wang Jiaqi and Yu Haixia realized multi-angle observation of equipment through ray casting and other methods based on the Unity engine[iii]. Chen Jia et al. integrated BIM and digital twin technologies to build a full-life-cycle management platform[iv]. The research team led by Chen Xiangzhang developed a virtual training system for roadheaders[v], and Guo Ning designed an AI training center, both of which integrated multiple technologies to improve training effectiveness[vi]. The Ministry of Education has issued policies to promote development: it selected 300 national-level virtual simulation experimental teaching centers during 2013–2015, and formulated a plan in 2017 to select 1,000 demonstration projects by 2024. Many domestic universities have responded actively. Institutions such as China Agricultural University and the University of Science and Technology of China have built relevant virtual simulation platforms respectively, and Shandong Agricultural University has also integrated such platforms with the rail transit system. However, domestic research still has several shortcomings. For example, at the technical integration level, voice recognition has insufficient adaptability to rail transit professional terminology, and the real-time performance of complex gesture recognition needs to be improved. In terms of interaction methods, the existing teaching resources have a single mode of interactive input and control, which cannot meet the diverse learning needs.

Foreign countries started early and have a wide range of applications in the field of rail transit virtual simulation, focusing mainly on two directions: First, the reconstruction of teaching scenarios. A complete teaching system with rail transit characteristics has not yet been formed, and research on the adaptability of technologies in different teaching scenarios remains weak. Second, the optimization of

International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

system operation, where fruitful achievements have been made. For example, DneciPtru used virtual simulation to analyze the impact of information management on rail transportation systems[vii]; Tanaino et al. simulated freight scheduling to find optimal cost solutions[viii]; Zubkov et al. explored paths to improve the efficiency of rail-sea intermodal transportation[ix]; SUN and WANG evaluated the performance of Personal Rapid Transit (PRT) systems using virtual simulation[x]; and Doran developed the Rail Profit Model (RPM) to test speed strategies for rail transportation[xi]. In addition, scholars have laid a certain foundation in the research and application of multimodal interaction. For instance, Ren et al. presented a comprehensive discussion on multimodal data fusion[xii]; Zhao analyzed the conceptual integration model of multimodal metaphor construction[xiii]; Tang carried out innovative applications of interactive virtual reality technology[xiv]; Tao et al. conducted a survey on multimodal human-computer interaction[xv]; Ye et al. conducted research on the construction of a precision teaching support system[xvi]; and Yang explored a new form of spatial construction from the perspective of embodied cognition[xvii]. However, several challenges persist[xviii]: In terms of teaching applications, although technological integration has been explored, the teaching system is still imperfect. In system management and decision optimization, simulation models lack sufficient accuracy in simulating complex systems and fail to respond in real time to dynamic factors. In the research of emerging technologies and energy conservation, virtual simulation is mostly used for preliminary scheme evaluation, and the feedback and correction mechanism between later-stage actual data and simulation results is not sound.

3. THEORETICAL FRAMEWORK

3.1 Core Theoretical Cornerstones

Traditional training methodologies are largely grounded in the theory of "disembodied cognition," which posits that learning is a process of the brain processing abstract data symbols and information, while physical and sensory experiences are marginalized. This orientation results in issues such as overemphasizing theoretical explanation over practical perception and prioritizing rote memorization of procedures over situational adaptation. The theoretical framework of this study shifts the training paradigm toward "embodied cognition" and "situated learning." Directly addressing traditional pedagogical pain points—such as lack of intuition, poor operability, limited hands-on opportunities in field settings, and the inability to reproduce real environments—this research constructs 1:1 high-fidelity 3D virtual scene. These scenes serve as digital physical carriers for "embodiment" and "context." Consequently, students are no longer mere observers; they can "enter" and "operate" within these contexts through diverse interaction modalities, thereby rebinding the cognitive process with experiential perception.

3.2 Three-Layer Architecture Model

The three-layer architecture, comprising the "Immersive Scene Layer, Multimodal Interaction Layer,

International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

and Cloud Data Intelligence Layer" based on embodied cognition and situated learning theories, constitutes an organically synergistic and hierarchically supportive system.As the core foundation and content carrier, the Immersive Scene Layer constructs a digital twin environment for students to "enter" and "explore," providing the physical space and objects essential for embodied cognitive experiences. The Multimodal Interaction Layer functions as the bridge connecting learners with virtual scenes and serves as the operational interface; it translates user intentions into precise control over virtual objects and provides dynamic feedback, facilitating the transition from "passive observation" to "active manipulation"—the critical link for "bodily involvement in cognition." The Cloud Data Intelligence Layer acts as the "brain" and resource hub, tasked with storing, managing, and scheduling all model data, pedagogical content, and learning records. It ensures cross-terminal consistency and real-time synchronization while supporting personalized learning and assessment through data analytics. These three layers are interconnected through a cycle of mutual empowerment, forming a closed loop of "environment construction, behavioral interaction, data feedback, and optimization iteration." This integrated architecture transcends the spatial, temporal, cost, and safety constraints of traditional training, creating a repeatable, low-cost, highly immersive, and evolving intelligent simulation model.

3.2.1 Immersive Scenario Layer

In this layer's construction, professional tools such as 3DS Max are utilized to precisely replicate every component of rail transit train control equipment according to actual proportions, ensuring that structure, dimensions, and kinematic relationships align with physical reality. A comprehensive environment—encompassing stations, sections, and control centers—is established to provide robust support for students' free exploration. Simultaneously, equipment operating principles, interlocking logic, and train operation rules are embedded into the virtual environment to ensure that operational feedback strictly adheres to real-world regulations.

By integrating 3D modeling and rendering technologies with game engines, the system presents more than just static models; it drives dynamic changes triggered by interaction. Abstract and complex internal structures and principles are transformed into visible, sectionable, and 360-degree observable 3D entities. By replacing expensive physical hardware with digital assets, a single system enables infinite reuse for a large student population, effectively resolving the contradiction between "limited equipment and high student demand." Furthermore, applying digital twin concepts creates a dynamic digital mapping of physical entities, ensuring virtual equipment serves not merely as visual models but as functional and state-based simulations.

(1) Hardware Sensing Interaction

1:1 scaled physical teaching aid models are fabricated via 3D printing, with embedded sensing

International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

technologies integrated into their interior—including Arduino development boards, sensors, joysticks, and alarms. Sensors capture users' operations on the physical models, while Bluetooth/Wi-Fi communication technologies transmit the physical operation data to the virtual scenario on the PC terminal in real time. Unlike simple external game controllers, these physical models are custom-developed for the shape and operational logic of professional equipment, ensuring that interaction actions are highly consistent with professional operations. This addresses the issue of "poor operability", delivers realistic force feedback and operational feel, and compensates for the lack of tactile feedback in pure mouse/keyboard-based interaction. It also trains muscle memory and operational proficiency. Combined with alarms, the physical models can issue real-time audio-visual alerts when incorrect operations occur (e.g., excessive force applied, wrong operation sequence), reinforcing the formation of correct operational patterns.

(2) Mobile Terminal (AR) Interaction

A mobile application will be developed to recognize real-world targets via the camera, overlay and display 3D models of equipment on the screens of mobile phones or tablets. This allows students to "hold" virtual equipment in their hands for repeated disassembly and observation of the linkage relationships of internal components, overcoming the cognitive barrier of physical equipment characterized by "high sealing performance and invisible internal structures". It extends learning from fixed training rooms to classrooms, libraries and dormitories, enabling "anytime, anywhere experimental training". The implementation mainly adopts Unity as the primary development platform, and completes augmented reality functions through the joint application of Unity and the Vuforia SDK development kit, achieving stable registration and tracking of virtual models in real-world images. Its basic architecture is shown in the figure below.

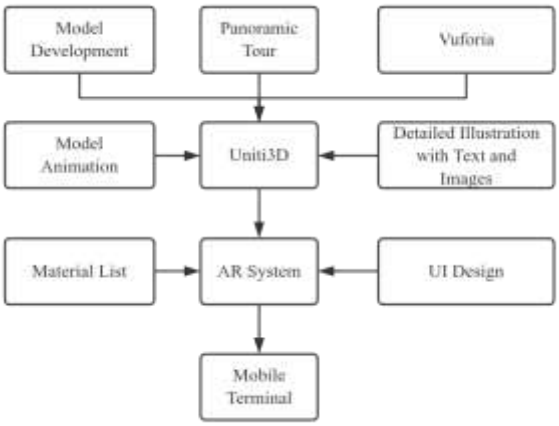

Figure 1 Unity Architecture

(3) Voice/Gesture Interaction

International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

Combining virtual reality technology with Kinect sensors, virtual simulation 3D models and scenarios, this module captures users' gestures and skeletal movements, as well as recognizes and parses their voice commands, thereby realizing Kinect-based human-machine interaction with equipment in virtual railway scenarios. It enables users to interact with the virtual environment through natural voice commands or specific gestures, issuing instructions via "speaking" and "gesturing"—the same collaborative methods used with peers in the real world. This further blurs the boundary between the virtual and real worlds and enhances the sense of presence. Meanwhile, it supports combined "voice + gesture" commands (e.g., saying "Move this to that place" while pointing to the target with a hand), resulting in a richer and more intelligent interaction dimension.

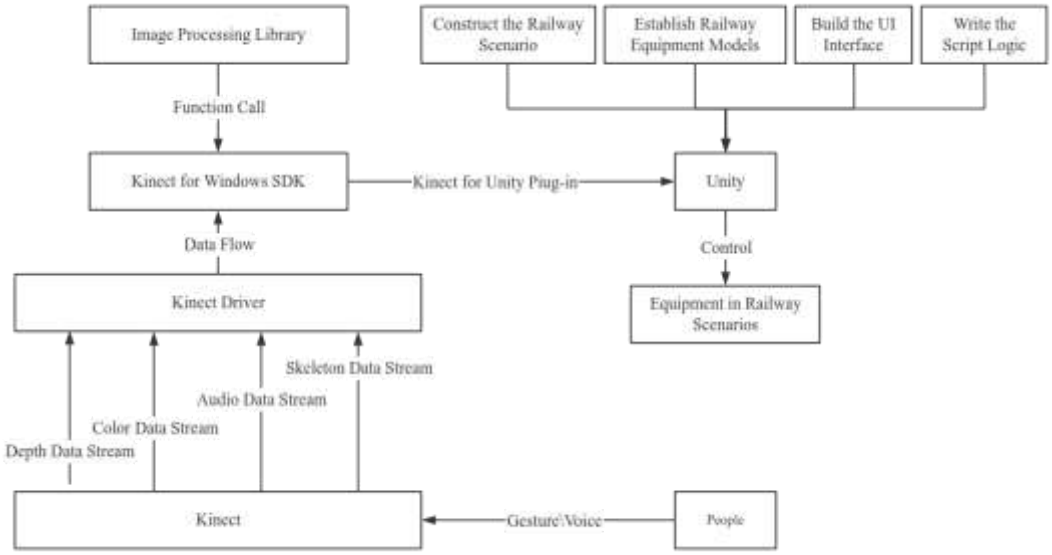

Figure 2 Human-Computer Interaction Structure Diagram

(4) Remote Holographic Interaction

Leveraging the rapid development of 5G networks, this module enables 1:1 real-time projection of teachers' life-size images from remote locations into local classrooms via holographic projection technology, realizing face-to-face remote interactive teaching. It mainly adopts holographic projection and capture technologies: green screens and high-definition cameras are used for teacher image capture at the instructor end, while holographic projection devices are deployed to display stereoscopic images at the student end. This fundamentally addresses the problems of "insufficient teaching faculty and unbalanced distribution of educational resources", and enables cross-regional multi-person collaborative teaching and hands-on practice. Students in different classrooms can collaborate or debate around the same virtual scenario, allowing learners from remote areas or institutions with weak faculty strength to receive real-time guidance from top experts and thus promoting educational equity.

International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

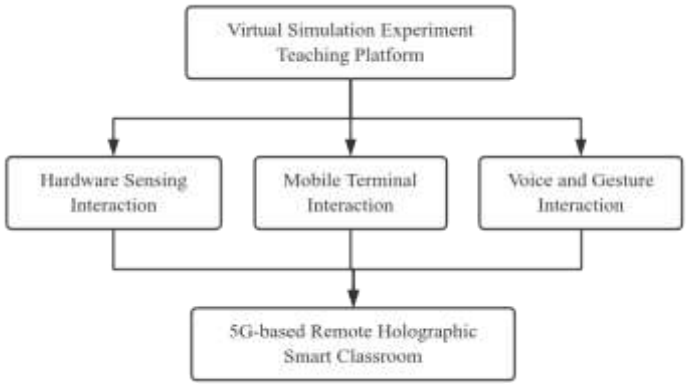

Figure 3 Overall System Functional Architecture

3.2.3 Cloud Data Intelligence Layer

The cloud server assumes the critical responsibility of storing all 3D models, scene data, course content, and user profiles, ensuring that content accessed across all terminals—including PCs, mobile devices, and holographic classrooms—is consistent and up-to-date. Through a cloud-based update mechanism, new training cases and fault scenarios are continuously injected, keeping the system current and preventing it from becoming a "one-off" product. Simultaneously, by maintaining detailed records of learning data, instructors can pinpoint individual weaknesses to provide differentiated guidance.The training platform is no longer a standalone simulation software, but a sustainable and evolving online educational service platform. This enables teaching administrators to manage virtual training content and assessments with the same efficiency as managing online course platforms.

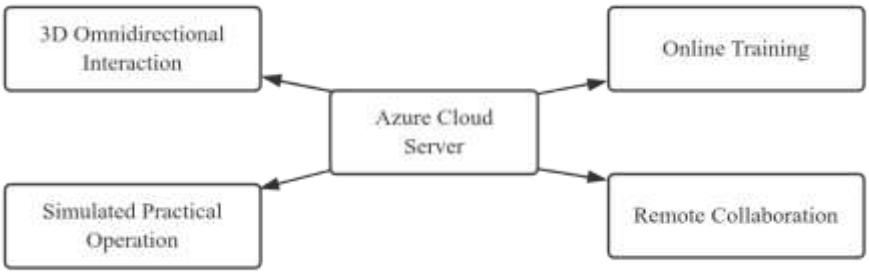

Figure 4 Cloud Server Architecture

This three-tier architecture establishes a high-fidelity, remotely accessible rail transit virtual training environment, effectively addressing core pain points in traditional training such as high equipment costs, operational risks, and unbalanced teaching resources. Through immersive scenes and multi-modal interactions, the model provides students with intuitive, repeatable, and low-risk hands-on training, significantly enhancing their practical skills and safety awareness. Meanwhile, the cloud-

International Journal of Research in Commerce and Management Studies

ISSN 2582-2292

Vol. 8, No. 02 March-April; 2026 Page. No. 477-486

based data intelligence layer provides robust support for centralized management and dynamic updates of resources. This enables wide sharing of premium teaching materials and personalized learning guidance, offering an innovative and efficient solution for the large-scale, high-quality cultivation of professionals in the rail transit field.

4. CONCLUSION

Based on the urgent demand for practical skills and safety literacy in the rail transit industry, this study focuses on prominent pain points in traditional training, such as equipment scarcity, high operational risks, and limited interaction methods. Grounded in embodied cognition and situated learning theories, we have developed a three-dimensional theoretical framework and training system solution comprising an "Immersive Scene Layer," a "Multi-modal Interaction Layer," and a "Cloud Data Intelligence Layer."

At the technical implementation level, the system deeply integrates cutting-edge technologies—including 3D modeling, Augmented Reality (AR), Kinect sensing, and 5G holographic projection—to establish four interaction modalities: hardware-perceptual, mobile AR, voice/gesture, and remote holographic interaction. On one hand, the system precisely replicates real operational tactile feedback and equipment logic through custom 3D-printed teaching aids and digital twin technology. On the other hand, it breaks spatial and temporal constraints via a mobile app, enhancing the sense of presence through natural interaction. Simultaneously, the cloud server enables resource sharing and dynamic updates, effectively resolving core contradictions in traditional training such as "limited equipment, poor operability, invisible internal structures, and uneven faculty distribution."

In terms of application value, the system focuses on rail transit train control equipment, transforming abstract principles and complex interlocking logic into visual, operable, and repeatable virtual training scenarios. This transformation not only reduces training costs but also strengthens students' operational proficiency and safety awareness. It provides an efficient, safe, and flexible new pedagogical tool for professional talent cultivation in rail transit, while offering a replicable practical paradigm for the deep integration of virtual simulation technology in engineering education.

Looking ahead, driven by continuous technological advancement and growing educational demands, this research will leverage its unique advantages to spark further innovation and transformation across the educational sector. We remain committed to keeping pace with the times, consistently exploring and practicing new methodologies to meet increasingly diversified educational needs and contributing to the cultivation of high-quality professional talents.

ACKNOWLEDGMENTS

This research is funded by the university-level "Innovation and Entrepreneurship Training Program for College Students" of Tianjin University of Technology (202510060079).

REFERENCES

[1] Yu J, Wang L, Zhang W, Zhou C. Winning markets via live-streams: Competitive manufacturers' channel strategies[J]. Journal of Retailing and Consumer Services, 2025,87:104391.

[2] Xie R. Urban Rail Transit Control and Simulation Based on Distributed Virtual Reality Technology[J]. Yangtze River Information & Communications, 2023, 36(04): 157-160.

[3] Wang J, Yu H. Implementation of Virtual Interaction in 3D Panoramic Roaming System from the Perspective of Virtual Reality[J]. Technology Innovation and Application, 2023(34): 28-31.

[4] Chen J, Wang J, Chen X. Full Life Cycle Management of Rail Transit Under Digital Transformation: Collaborative Application of BIM Technology and Digital Twin[J]. Public Transport of China, 2024(22): 76-78.

[5] Chen X, Yin Z, Wang Z. Application of Virtual Training Technology in Industrial Technology Training[J]. Experimental Technology and Management, 2015, 32(05): 127-131.

[6] Guo N, Yang D, Yan G. Design and Implementation of an Open Smart Training Center for Urban Rail Transit Vehicles[J]. Urban Rail Transit Research, 2025, 28(03): 108-111.

[7] DneciPtru, D. Technology Information Management Applied To Rail Transportation System[J]. Ovidius University Annals: Economic Sciences Series, 2020, 20(1): 598-603.

[8] Tanaino, I., Yugrina, O., Zharikova, L. Assessment criteria for decisions in the field of rail freight transportation[C]. EDP Sciences, 2018: 02015. In: MATEC Web of Conferences, Vol. 216.

[9] Zubkov, V, Sirina, N, Popovic, Z, Manakov, A, Breskich, V. Improvement of Cargo Transportation Technology in Rail and Sea Traffic[C]. TransSiberia 2019.Advances in Intelligent Systems and Computing, vol. 1116. Cham: Springer, 2020: 1110-1119.

[10] Sun, S, Wang, B. Low-energy Mountain Transportation System with PRT Rail Transit Technology[J]. Journal of Landscape Research, 2020, (3): 15-17, 26.

[11] Doran, M. P. Profit based simulation model for the rail transportation industry[D]. Norfolk, VA: Old Dominion University, 2016.

[12] Ren ZY, Wang ZC, Ke ZW, et al. A Review of Multimodal Data Fusion[J]. Journal of Computer Engineering & Applications, 2021, 18: 49.

[13] Zhao XF. The conceptual integration model of multimodal metaphor construction:a case study of a political cartoon[J]. Foreign Languages Research, 2013: 9.

[14] Tang T. Interactive virtual reality technology empowers the digital inheritance and innovation of intangible cultural heritage[J]. 10.12677/isl.2025.94071.

[15] Tao JH, Wu YC, Yu C et al. A Survey on Multi-modal Human-Computer Interaction[J]. Journal

of Image and Graphics, 2022: 1697.

[16] Ye XD, Liu ZM. Research on the Construction of Precision Teaching Support System Based on Multimodal Large Models[J]. Journal of Distance Education, 2024, 1: 84.

[17] Yang. The Interaction of Virtual and Real: Exploring a New Spatial Construction from the Perspective of Embodied Cognition[J]. New Architecture, 2022, 6: 87.

[18] Zhou C, Bai D, Li T, Yu J. Personalized recommendation, behavior-based pricing, or both? Examining privacy concerns from a cost perspective[J]. Omega, 2025, 133: 103223.