MIFE: A Multimodal VR Immersive Training System for Fire Escape

Weijie Liu[*]
Tianjin University

Yalei Liu[†]
Tianjin Yunlan Internet of Things Technology Co., Ltd.

Jiaxuan Gao[‡]
Tianjin University

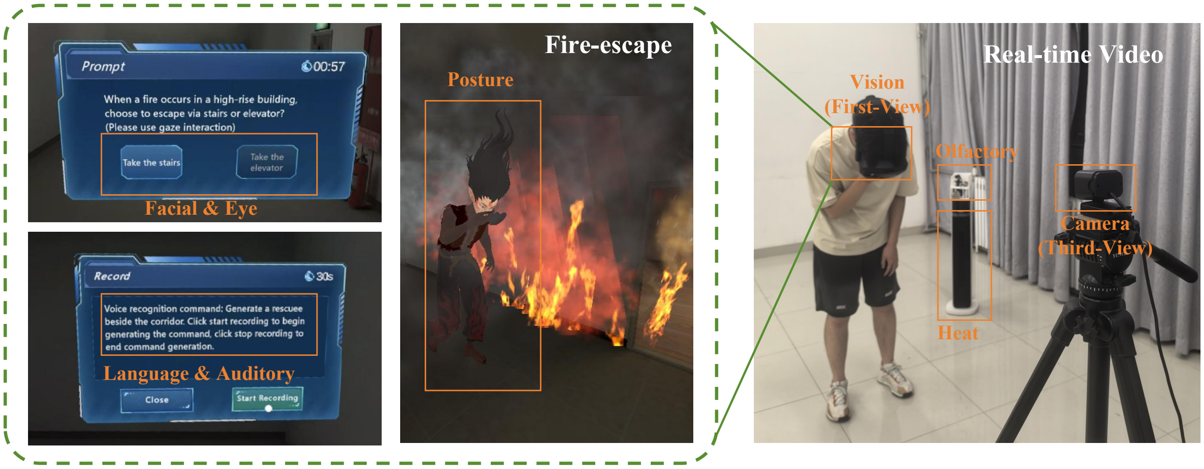
Zixuan Xie[§]
Tianjin University

Qiuyu Fu[¶]
Tianjin University

Lu Lu[‖]
Tianjin University

Jian Ma[**]
Tianjin University

Kun Li[††]
Tianjin University



When a fire occurs in a high-rise building, choose to escape via stairs or elevator? (Please use gaze interaction)

Voice recognition command: Generate a rescuee beside the corridor. Click start recording to begin generating the command, click stop recording to end command generation.

Figure 1: MIFE integrates 8 modalities, including pose, auditory, heat, eye tracking, facial expression, vision, olfactory, and language. Among them, the posture adopts dual-perspective motion capture with an RGB camera. This method maps the trainees' full-body posture to the virtual avatar in real-time, improving the trainees' skill acquisition. We conducted user study to evaluate the effectiveness and usability of the system.

Abstract

High-rise fire escape training for the public poses a significant challenge, owing to the prohibitive cost and logistical intricacies involved in deploying high-fidelity simulations. While virtual reality (VR) technologies have demonstrated potential in safety education, the number of interaction modalities they offer is mostly limited to 3 to 5 types, and they heavily rely on controllers that abstract actions into button presses, limiting immersion and skill transfer. In this paper, we propose the MIFE system, which designs 8 distinct interaction modalities along with a real-time dynamic fire spread simulation to further improve immersion, allowing trainees to perceive the fire scene and efficiently master key escape skills through vision, auditory, olfactory, etc. Moreover, we have designed a controller-free and real-time full-body motion capture (MoCap) module, achieving precise mapping between the trainees' full-body movements and the virtual avatar. Additionally, MIFE leverages a knowledge graph to provide tailored guidance, adapting dynamically to trainees with varying levels of fire escape proficiency. The results of the user study demonstrate that MIFE significantly outperforms self-study and controller-based training systems, particularly, improving by 3.08/10 and 2.38/10 in the performance score of the evaluation stage, respectively. This implies its practical utility and potential for broader adoption in high-rise fire emergency training.

Index Terms: Fire Escape, Virtual Reality, Real-time Full-body Motion Capture, Training System.

1 Introduction

High-rise building fires pose a severe global public safety threat, making effective fire safety training an essential line of defense for saving lives and reducing casualties. Global statistics indicate that the annual incidence of residential fires ranges from 8 to 20 cases per 100,000 population [1], with more than 400 people dying in fires and 19,000 people being injured every day around the world [2]. Fire escape skills, thus, constitute critical emergency competencies that can substantially enhance chances of survival through three key measures: scientific escape route planning, effective smoke prevention strategies, and proper application of self-rescue and mutual-aid techniques.

Traditional fire escape training typically relies on teacher-guided instructions through video supplementation [3, 4], an approach that suffers from insufficient hands-on practice and a lack of personalization. While on-site demonstrations offer greater practicality, they require specialized instructors and dedicated facilities, thereby imposing organizational challenges and limiting the number of trainees. Therefore, these constraints reduce training flexibility and scalability, ultimately hindering widespread adoption.

Although VR-based fire escape training has emerged as a promising alternative due to its immersive capabilities, immersion in fire escape training remains challenging. While some methods

[*] e-mail: wjliu@tju.edu.cn. Contributed equally.
[†] e-mail: liu_yl2022@tju.edu.cn. Contributed equally.
[‡] e-mail: 3023244275@tju.edu.cn
[§] e-mail: 2025244155@tju.edu.cn
[¶] e-mail: qyfu219@tju.edu.cn
[‖] e-mail: lulu_998543@tju.edu.cn
[**] e-mail: jianma@tju.edu.cn
[††] e-mail: lik@tju.edu.cn. Corresponding Author.

[5, 6, 7, 8] adopt multiple modalities—such as visual, auditory, tactile, and thermal —these approaches still face limitations in fully replicating real-world fire scenarios [9]. A key limitation arises, for instance, when trainees notice something unusual via olfaction, they probably take action immediately rather than confirm this via vision. This case implies that current multi-modality-based approaches hardly enable trainees to form stress instincts and make rapid, correct decisions in actual emergencies. Moreover, most systems [10, 11, 12] relying on controllers ignore critical actions, such as grasping a towel or bending down to cover someone's nose, to abstract button inputs, resulting in a significant discrepancy between simulated operations and real-life escape behaviors. This mismatch violates the principle of encoding specificity [13, 14], thereby undermining the transferability of training effects to actual emergency scenarios [15]. Furthermore, existing systems [16, 17] fail to include real-time fire spread and personalized adjustments for trainees, making them difficult to strengthen training content tailored to trainees' unfamiliar knowledge [18].

Hence, the above-mentioned shortcomings highlight the need for advanced VR systems that improves immersion, real-to-virtual motion synchronization and dynamic adaptability to ensure training transfer.

To address these challenges, we present a novel intelligent immersive VR system, MIFE, for fire escape training. The key idea is to cover more modalities, create a more immersive training platform that captures trainee motions in real-time using an RGB camera, and dynamically adapt to different trainees. For constructing immersive experiences, MIFE achieves realism through the integration of eight sensory modalities: posture, auditory, thermal, eye tracking, facial expression, vision, olfactory, and language, creating a truly multisensory simulation environment. To eliminate reliance on VR controllers and ensure alignment between trainee actions and real-world escape maneuvers, we develop a real-time motion capture system that integrates both third-view and first-view. This new approach designs a bidirectional motion fusion mechanism and adaptive viewpoint switching technology, which dynamically optimizes perspective rendering based on trainee posture and environmental interaction requirements. To enhance dynamic adaptability, a real-time fire spread model based on cellular automata simulates the evolution process and spread behavior of fire. A large language model based on a knowledge graph delivers fire knowledge according to trainee basic information and scene operation performance.

In summary, our key contributions are as follows:

- We propose a fire escape training system, MIFE, integrating eight multimodal data streams - pose, auditory, heat, eye tracking, facial expression, vision, olfactory, language - with our self-developed olfactory divergent device to further enhance immersive experience.

- We employ full-body dual-perspective motion capture. The system provides precise whole-body posture data and adaptive viewpoint switching, offering a transferable solution for real-world fire escape scenarios.

- We implement a real-time fire spread simulation using cellular automata and construct a safety education knowledge graph for fine-tuning large language models, enabling dynamic adaptation in our training system.

- We demonstrate the system's effectiveness in both learning outcomes and trainees experience through user studies and ablation. By providing an immersive learning environment with realistic operational experiences, it facilitates knowledge transfer.

2 Related Work

2.1 Traditional Fire Safety Training

Traditional fire training presents a trajectory evolving from passive knowledge dissemination to active, hands-on exploration. Theoretical instruction, represented by classroom teaching, lectures, and multimedia, is widely adopted. For example, Chavez et al. [3] use multimedia training to provide knowledge and skill education for children and parents of different age groups, while Lee et al. [4] employ video-based training for general and scenario-specific fire safety education for medical personnel. However, this unidirectional information delivery, due to its lack of interactivity, hinders deep engagement of trainees and the effective internalization of knowledge.

To overcome these limitations, researchers introduce interactive elements like tabletop simulations to improve engagement and memory retention, such as the work by Delcea and Cotfas [19] which trains students' decision-making during evacuations. As these methods lack the situational psychological pressure of a real fire, live drills are introduced. Studies like Najmanová et al. [20] demonstrate that drills significantly reduce escape problems, and Huang et al. [21] propose a live-fire training system. Nevertheless, this approach is constrained by high costs, organizational difficulty, and insufficient safety, limiting its feasibility as a regular, large-scale training tool. It is against this backdrop that digital, interactive solutions like serious games have emerged, aiming to provide more engaging learning experiences in a safe and cost-effective manner.

2.2 VR-based Fire Safety Training

The emerging paradigm of VR-based emergency training offers a systematic and reliable solution for fire safety [22, 23]. The interactive modes of VR bridge the crucial gap between passive knowledge acquisition and active skill application [24, 25]. By immersing trainees in realistic 3D environments, VR also replicates the intense psychological pressure of real fire scenarios [26]. Critically, it provides a zero-risk, low-cost setting for repeatable practice, allowing trainees to learn safely from their mistakes [27, 28].

Research shows that a strong sense of presence, induced by high immersion, significantly improves fire emergency training outcomes, particularly for escape drills [22]. We have compared core literature from the last three years in Tab. 1. Despite this progress, a significant gap remains between VR's potential and the effectiveness of current systems. Some researchers prioritize enhancing immersion with hardware, but this commonly requires cumbersome equipment. For example, Yang et al. [29] utilize thermal suits and motion platforms to create an immersive fire environment. Others focus on environmental fidelity. For instance, Ling et al. [7] use a thermal box to simulate the high temperatures of a fire scene, or Narciso et al. [30] add olfactory cues with a SensoryCo SmX-4D display. These efforts, however, typically result in a sparse combination of partial modalities, failing to achieve effective training transfer through a holistic, full-dimensional modal architecture.

Interaction in fire escape training systems is commonly controller-based [7, 10, 11]. To enhance immersion and skill transfer, recent work has begun aligning trainee and avatar movements. For instance, Kang et al. [5] employ pose trackers to achieve partial body adaptation. Nevertheless, such solutions have not yet eliminated the core dependency on controllers. This reliance imposes fundamental limitations on the natural mapping between real-world actions and virtual representations.

Regarding dynamic adaptability, many studies employ offline data from the Fire Dynamics Simulator (FDS) [16, 17]; This approach inherently restricts real-time interaction between trainee and environment. Meanwhile, real-time solutions, such as the state-machine-based model from Qazi et al. [6] or the cellular automata simulation by Bo et al. [31], sacrifice physical fidelity in complex scenarios by abstracting either propagation rules or 3D geometry.

Table 1: A comparison of related works on fire emergency training.

Reference	Number of Modalities	Full-body MoCap	Interaction Mode	Dynamic Fire Spread	Knowledge Graph-based Recommendation
Ling et al. [7]	4 (Vision, Auditory, Heat, Physiology)	✗	VR Controller	✗	✗
Hamed-Ahmed et al. [11]	4 (Vision, Auditory, Eye tracking, Facial expression)	✗	VR Controller	✗	✗
Oliveira et al. [8]	5 (Vision, Auditory, Tactile, Heat, Physiology)	✗	VR Controller	✗	✗
Kang et al. [5]	3 (Vision, Auditory, Tactile)	✗	VR Controller, Tracker	✓	✗
Yang et al. [29]	4 (Vision, Auditory, Heat, Tactile)	✓	MoCap Platform (Three-axis Controlled Motion Platform)	✓	✗
Qazi et al. [6]	3 (Vision, Auditory, Tactile)	✗	VR Controller	✓	✗
Ours	**8 (Pose, Auditory, Heat, Eye tracking, Facial expression, Vision, Olfactory, Language)**	✓	**Lightweight MoCap (Monocular Camera)**	✓	✓

Hence, in this paper, we introduce MIFE, a fire escape training system based on high-fidelity, full-body motion capture. By integrating photorealistic virtual scenes, multisensory feedback mechanisms, and precise synchronous mapping between trainee movements and the virtual avatar, the system achieves a faithful reproduction of the entire fire escape process.

3 System Design

We propose an immersive VR-based fire escape system, MIFE, with the framework structure shown in Fig. 2. Our goal is to create an immersive training platform that improves immersion, real-to-virtual motion synchronization, and personalized adaptation.

In Sec. 3.1, we introduce our synchronized multimodal interaction module, which includes posture, auditory, thermal, eye tracking, facial expression, visual, olfactory, and linguistic modalities. Our posture modality utilizes real-time dual-perspective motion capture (Sec. 3.2) to ensure the consistency between the trainee's actual posture and that of the virtual avatar. The adaptability of the system is achieved through the fire spread model and personalized knowledge recommendation: the fire spread model forms a realistic fire spread situation (Sec. 3.3), while the personalized knowledge recommendation module utilizes trainee personal information and operational performance (Sec. 3.4) .

3.1 Synchronized Multimodal Interaction

Multimodality enhances the experience in VR, improves overall performance, and generates capabilities in skill and knowledge transfer [32]. To enhance immersive and authentic experiences, the MIFE system designs 8 interaction modalities (measurements: pose, eye-tracking, facial expression, voice; stimuli: visual, auditory, olfactory, thermal) distinguishing it from other systems that typically utilize only three to five modalities. By stimulating multiple sensory dimensions in synergy, the system engages a broader spectrum of human perception, thereby providing trainees with a more realistic sense of immersion and authenticity.

Posture: MIFE employs a novel dual-perspective motion capture module that achieves real-to-virtual motion synchronization between trainees' natural movements and their virtual avatars, eliminating reliance on traditional VR controllers. Detailed information for dual-perspective motion capture can be found in Sec. 3.2.

Based on dual-perspective motion capture, trainees can control the avatar's forward movement and standing posture by lifting the right leg, trigger a 45° rotation to the right or left by raising the corresponding thumb, and interact with the UI through a pinch gesture. All other movements remain consistent with the trainee's actual motions.

Eye tracking: Eye tracking is considered an efficient interaction method in VR systems [33]. Hence, it is employed in our system for selecting key escape routes within the VR scene. Although physical interaction with objects cannot be achieved solely through gaze in actual fires, eye tracking, being the interaction mode closest to real-world observation, is used to interact with UI elements. This design forces trainees to actively complete hazard judgments, including observation, analysis, and decision-making.

Eye tracking interaction system is based on the built-in eye tracking module of the HTC VIVE Focus Vision headset, with a nominal accuracy of better than 1.1° and a sampling rate of 120 Hz. The system obtains the coordinates of the fixation points of both eyes through the Wave SDK. An interaction is registered as successful if the gaze remains on an interactive button for more than one second; otherwise, it is deemed invalid.

Facial expression: Facial expression tracking monitors the physiological and emotional changes of trainees. Once expressions such as fear and extreme pain are recognized, MIFE can exit in a timely manner to protect the psychological state of trainees.

Facial expression tracking system is based on the VIVE Focus series of facial trackers with a sampling rate of 60 Hz. The system uses Wave SDK to capture 38 mixed shapes of lips, teeth, tongue, cheeks, and chin to capture facial activities and analyze trainees' psychological emotions.

Language: MIFE enables generating a rescuee and dialogue with large language models through natural language interaction. Generating a rescued person can train trainees' ability to save others, and conversing with large language models can promptly resolve trainees' doubts. By harnessing natural language interaction, MIFE simulates emergency rescue scenarios, thereby enhancing training immersion.

This module adopts a three-layer architecture design: automatic speech recognition layer, natural language understanding layer, and text-to-speech layer. The automatic speech recognition layer employs a large ASR model [34] to transcribe input audio into a text sequence, Y. The Natural language understanding layer determines whether Y contains the keyword "generate". If the keyword is present, MIFE executes the rescued person generation; otherwise, it retrieves answer text via the large language model and performs text-to-speech synthesis.

Vision: The system complies with International Organization for Standardization (ISO) standards, including ISO 21542:2021 [35], ISO 30061:2007 [36], and ISO 16069:2017 [37], to accurately simulate real-world fire safety facilities in residential buildings in a virtual environment. These facilities include fire-rated self-closing doors, emergency lighting systems that meet safety standards, and properly positioned evacuation signage. This high-fidelity virtual training environment closely mirrors real-world scenarios, providing trainees with reliable fire evacuation training conditions.

Auditory: Sound, as a critical information carrier, can simulate real-world scenarios and enhance a sense of urgency [10, 7, 38]. However, excessive stimulation and distraction may hinder effective learning [39]. Therefore, in order to simulate the sound of a real fire while avoiding excessive stimulation, we use the 3D spatial

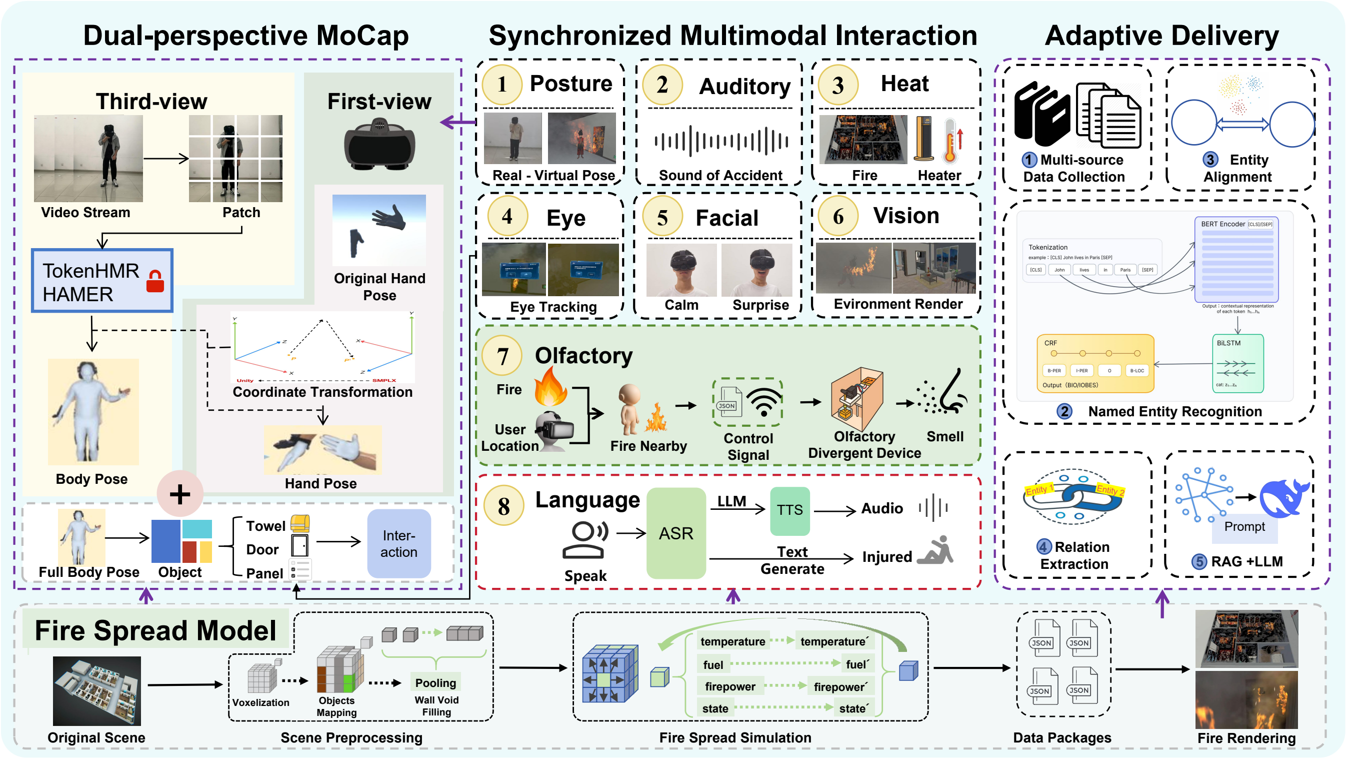


Figure 2: Overview of our immersive MIFE training system. (a) Dual-perspective MoCap: MIFE system captures full-body poses via third-person and hand poses via first-person view, enabling real-time interactive feedback. (b) Synchronized multimodal interaction module: MIFE enhances immersive training experiences by providing eight multimodal sensory channels: posture, auditory, heat, eye tracking, facial, vision, olfactory, and language. (c) Adaptive delivery module: MIFE system utilizes a knowledge graph-enhanced LLM to personalize fire escape guidance based on user profiles and real-time performance data. (d) Real-time fire spread model: This module performs real-time fire spread simulations, generating accurate spatiotemporal visualizations of flame diffusion in virtual environments.

audio effects in Unity to dynamically control the flame volume. As trainees move through the virtual environment, the sound dynamically adjusts its volume based on the distance and orientation from the fire source. This spatially dynamic audio feedback enhances the immersive experience and scene realism in virtual training.

Olfactory: Olfactory feedback simulates the smell of smoke and burning objects in a fire scene. The system calculates odor concentration values based on the trainee's location (see Sec. 2.1 in supplementary material), and delivers corresponding olfactory feedback accordingly, providing trainees with odor simulation experience.

MIFE transmits odor concentration data to the olfactory divergence device in JSON format through a TCP socket. The device receives continuous intensity values and linearly maps them to 8-bit PWM duty cycles in the range of 0-255, thereby synchronously adjusting the fan speed and servo motor trigger frequency.

Heat: Inspired by Ling et al. [7], we introduce a temperature feedback mechanism to further enhance trainees' immersion and authentic experience. By simulating the heat changes in a fire environment, trainees can more intuitively perceive the dynamic evolution of the temperature field, thereby enhancing their understanding of fire scenarios and emergency response capabilities.

The thermal feedback system operates a local area network. It controls the Philips 3000 A1 series tower heater through the Mi Home local communication protocol. The system obtains real-time temperature data (see Sec. 2.2 in supplementary material), linearly maps this data to a range of 15°C to 45°C, formats it into API commands in JSON format, and sends them to the heater to achieve dynamic temperature control of the equipment.

In terms of knowledge acquisition, trainees can map their movements onto their avatars through posture tracking, thereby internalizing correct knowledge of escape behavior. Eye tracking simulates natural visual attention and path selection mechanisms, thereby enabling trainees to acquire knowledge for directional decision-making. Voice interaction solves trainees' doubts and assists them in understanding key information. Multi-sensory feedback - visual, auditory, thermal, and olfactory - collectively provides critical scene cognitive knowledge. Facial expression recognition monitors psychological states and terminates training when extreme fear is detected to ensure safety.

3.2 Real-time Dual-perspective Motion Capture

Accurate posture tracking and hand tracking are crucial for fire escape training, as they enable high physical fidelity, facilitate effective skill transfer, and support action-based assessment and feedback. To ensure consistency between the trainee's posture and the virtual avatar's, we employ a monocular camera for third-person, full-body motion capture and an HMD for first-person hand tracking. This approach contrasts with methods that rely on VR controllers or XR headsets for hand tracking alone, which lack full-body information.

Third-person view: To acquire full-body human poses in real-time, our method employs the TokenHMR [40] model and the HAMER [41] model. Specifically, a monocular camera captures an image, I, in real-time. The input image I is processed by two models in parallel: the TokenHMR model outputs the SMPL [42] parameters for body pose θ_{SMPL}, and the HAMER model outputs the MANO [43] parameters for hand pose θ_{MANO}.

First-person view: In real-time hand tracking based on a third-

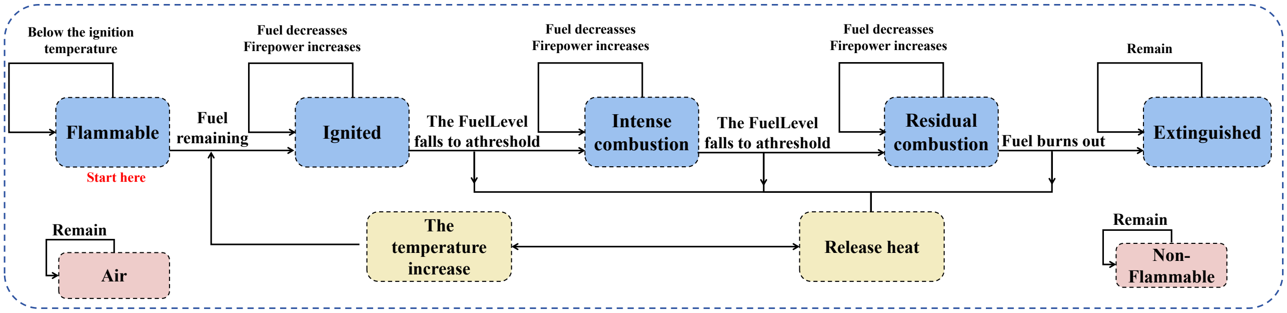

Below the ignition temperature
Flammable
Start here
Fuel remaining
Fuel decreases Firepower increases
Ignited
The FuelLevel falls to athreshold
Fuel decreases Firepower increases
Intense combustion
The FuelLevel falls to athreshold
Fuel decreases Firepower increases
Residual combustion
Fuel burns out
Remain
Extinguished
Remain
Air
The temperature increase
Release heat
Remain
Non-Flammable

Figure 3: State management and fire spreading model for dynamic fire simulation in real-time 3D environments.

person view, the hand may occupy a relatively small proportion of the camera's field of view, which can lead to deviations in the reconstructed pose. To address this issue, MIFE incorporates first-person view hand tracking for compensation. The HTC VIVE Focus Vision can provide first-view hand tracking data. Since Unity uses a standard left-handed coordinate system while SMPLX uses a non-standard coordinate system, the mapping involves a 90° clockwise rotation around the y-axis. At the same time, the skeletal structure of the hand model in Focus Vision does not align with that of the SMPLX human body model. Therefore, further processing is required to handle the rotational relationship with parent nodes, which is defined as follows:

$$R'_h = \begin{cases} R_{\mathrm{p}}^{-1} \cdot R_c, & \text{with parent node} \\ R_c, & \text{without parent node} \end{cases}, \tag{1}$$

here, R_{p}^{-1} is rotation of parent nodes and R'_h is final rotation of SMPLX. Finally, to ensure that there are no sudden changes in posture, hand posture data is processed using linear smoothing.

Fusion: First-view hand tracking can become unstable if hands are positioned outside the optimal range of the headset's built-in sensors, such as when they are too low or too high. In such cases, the system automatically switches to third-view hand tracking to guarantee stability and smooth motion for the virtual avatar. The transition is formally defined as follows:

$$P = \{\theta_{\mathrm{SMPL}}\} + \varepsilon R'_{\mathrm{h}} + (1-\varepsilon)\,(\theta_{\mathrm{MANO}})\,, \tag{2}$$

where θ_{SMPL} and β_{SMPL} are body pose and body shape from TokenHMR, θ_{MANO} is hand pose from HAMER and ε is fusion weight, which is 0 or 1.

In summary, MIFE employs a distributed architecture that integrates third-view full-body motion capture with first-view hand tracking, achieving real-time precise mapping between trainees' physical movements and virtual avatar motions. By eliminating reliance on traditional VR controllers, this fusion solution overcomes occlusion and limited field-of-view challenges inherent in single-view motion capture systems, enhancing the reliability and naturalism of full-body motion interaction.

3.3 Real-time Fire Spread Model

In order to enable trainees to perceive dynamic changes in fire scenarios, we employ a cellular automata framework to construct a model for fire propagation. Unlike the approach using FDS, our approach does not require lengthy computation times and significantly reduces the amount of data generated. It also differs from experience-based real-time methods, as those methods often lack consistency with real fire spread in high-rise buildings.

Inspired by Byari et al. [44], we develop a three-dimensional cellular automaton model based on combustion heat and conduction heat. To this end, the model is structured around a global temperature field and defines a seven-dimensional state space, S, to represent the dynamics within complex high-rise environments. The state transition is shown in Fig. 3. The state transition depends on the current state s and the temperature at the next moment T^{t+1}. T^{t+1} is defined as:

$$T^{t+1} = T^t + \sum_{i \in N} \beta \cdot \kappa \cdot \left(T_i^t - T^t\right) + \Delta T, \tag{3}$$

where ΔT represents the heat released by combustion, β is the thermal conduction regulation coefficient controlling the heat absorption/release ratio of cells, and κ is the material's thermal conductivity coefficient. T_i^t denotes the temperature of neighboring cell i at time step t, and T^t represents the temperature of the central cell at time step t. ΔT is defined as follows:

$$\Delta T = \begin{cases} -k \cdot \alpha \cdot (F^t - u)^2 + \dfrac{k \cdot \alpha \cdot u}{2}, & F^t \leq u \\ k \cdot \alpha \cdot (F^t - u)^2 + \dfrac{k \cdot \alpha \cdot u}{2}, & F^t > u \end{cases}, \tag{4}$$

where k is the combustion control coefficient, α denotes the material's burning rate, u represents the fuel threshold coefficient, and F^t indicates the current time of fuel value. The temporal evolution of F^t is specifically defined as follows:

$$F^t = F^{t-1} - \gamma \cdot \alpha \cdot \frac{T^{t-1}}{T_{\mathrm{ign}}} \cdot F^{t-1}, \tag{5}$$

where γ denotes fuel consumption coefficient, and T_{ign} represents material's ignition temperature. F^{t-1} is the last time of fuel value and T^{t-1} is the temperature of the central cell at the last time step.

In summary, this cellular automaton model focuses on the structural characteristics of high-rise buildings and fire combustion dynamics. By performing global tensor operations on the temperature field, this model constructs an efficient real-time dynamic fire spread model that maintains consistency with real fire spread.

3.4 Personalized Knowledge Recommendation

Our study first systematically collects high-quality popular science text data from authoritative books, clinical guidelines, and other reliable sources. We then apply a BERT-BiLSTM-CRF [45] model architecture to perform entity recognition in the domain of science popularization. Due to the diversity of data sources, the extracted public science knowledge inevitably contains duplicate information and redundant expressions. To address this, entity alignment is introduced as a crucial refinement step prior to constructing a multi-source knowledge graph. Specifically, we first conduct preliminary entity merging based on semantic similarity [46], followed by manual review and deduplication by two senior domain experts, ensuring that only core domain entities are retained. Subsequently, a relation extraction (RE) model [47] is employed to identify semantic relationships between entities, and high-quality triples containing core entities are preserved, resulting in a structured knowledge graph tailored for public science communication.

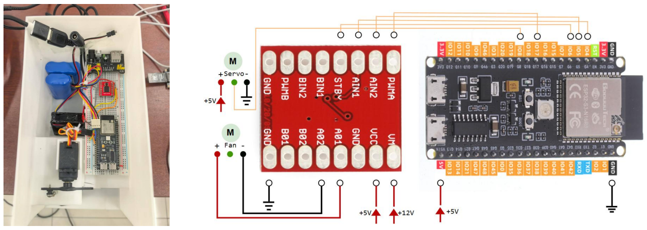
At the application level, this knowledge graph is integrated as an external knowledge base for DeepSeek [48], where it works in conjunction with a large language model to enable personalized learning content recommendation and precise learning report generation. These features are deployed in the “question recommendation” stages, allowing customized outputs based on learners’ age, occupation, and other background factors, thereby enhancing both training effectiveness and learning experience.

4 IMPLEMENTATION

4.1 Hardware Implementation

The system runs on an RTX-4070Ti Super GPU, an HTC Vive Focus Vision, a monocular camera (Newmine 4K), a Philips Tower Heater Series 3000 A1, and an olfactory divergent device.

Olfactory divergent device: The olfactory divergent device, as shown in Fig. 4 (a), is constructed using 3D printing and electronic components, including ESP32 chips, servo motors, etc. The circuit design is shown in Fig. 4 (b). The program running on the ESP32 chip is written and uploaded using Arduino IDE. Its main function is to connect to the host computer via a local area network and receive JSON-formatted commands that control the servo motor trigger frequency and fan operation parameters.



(a) Olfactory device

(b) The circuit design diagram

Figure 4: Specific implementation of the olfactory divergent device.

4.2 Software Implementation

The VR program of this system is developed on Unity (version: 2022.3.55f1c1) and is packaged to run on a headset, with data transmission through WebSocket.

Full-body motion capture: The system collects video streams through OpenCV on the PC, preprocesses them, and transmits them to the TensorRT engine on the GPU to synchronously infer human pose, shape, and hand pose. The inference results are transmitted in real-time to the VR headset through WebSocket, driving the movement of the virtual character.

Finally, full-body motion capture runs at approximately 30 FPS. The accuracy of our 3D pose estimation in Mean Vertex Error (MVE), Mean Per Joint Position Error (MPJPE), and Procrustes-Aligned Mean Per Joint Position Error (PA-MPJPE) on the EMDB [49] dataset is 104.2, 88.1, and 49.8, while on the 3DPW [50] dataset it is 86.0, 70.5, and 43.8, respectively.

Hand tracking: Hand tracking on the HTC VIVE Focus Vision uses the official Wave SDK for Android. The system captures real-time local joint rotations, applies sequential transformations to compute global rotations, and drives hand animation at approximately 45 FPS.

Fire spread model: The fire spread model is implemented based on PyTorch tensor to utilize GPU acceleration for large-scale parallel computing. In the input stage, the system voxels the 3D scene into a tensor mesh, where each voxel contains a dynamic state vector that records material properties, temperature, fuel mass, and discrete state identifiers (such as non-combustible, combustible, active combustion, or burned out). In diffusion simulation, the model integrates thermodynamic mechanisms and the global fire state field through vectorized Boolean masks to drive state transitions, thereby achieving the diffusion process of fire in voxel space.

Knowledge recommendation: The large language model component of the system is deployed on the Coze platform as multiple task dialogue agents, each implemented based on the Deepseek model. The backend uses FastAPI and Pydantic to build a lightweight orchestration layer, providing RESTful endpoints to receive and manage structured user context and VR training data. When the free question and answer task is triggered, the server serializes the relevant context into a JSON string in UTF-8 format and forwards it to the corresponding Coze dialogue robot through the official Coze Python SDK (cozepy). The front-end extracts the outermost JSON block from the received streaming response text and parses it into a structured object based on predefined patterns, completing rule-based data format validation. Finally, the average latency for LLM is 3424ms, and the average latency for STT and TTS is 100ms, with a total process time of 3624ms.

For more implementation details of each component, please refer to the supplementary materials.

5 USER STUDY

In order to verify the effectiveness of the MIFE system in fire escape training and compare it with self-study and controller-based training in the VR system, this study designed a controlled experiment, including subjective scale scoring, behavioral experiment measurement, and theoretical knowledge evaluation.

5.1 User Study Design

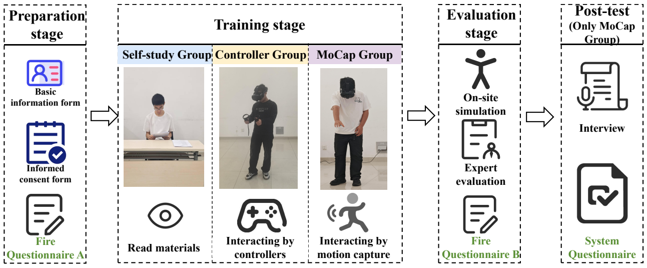


Figure 5: Experiment flow.

5.1.1 Participants

Prior to participant recruitment, the study protocol was reviewed and approved by the Institutional Review Board (IRB) of our university (Protocol Number: TJUE-2024-256; Ethics Committee: Tianjin University Ethics Committee). We recruited 40 adults (30 for user study, 10 for ablation; male: 23, female: 17), aged 19 to 32 years (mean: 22.53, SD: 3.19), to participate in the study and obtained the consent of all participants.

To mitigate the influence of prior fire escape proficiency and VR experience on the experimental outcomes, we adhered to the participant selection criteria: (1) participants had no prior experience with VR; (2) participants had not undergone fire escape training previously; (3) participants exhibited no cognitive or physical impairments that would hinder their ability to perform fire escape.

5.1.2 Procedure

Experimental preparation stage: Participants signed an informed consent form, registered their basic information, and completed a fire safety questionnaire A as a basic test to determine their respective levels of fire safety knowledge. Finally, participants randomly selected an envelope to determine their assigned group. 30 participants were randomly divided into 3 groups on average: self-study group, controller group, and MoCap group.

- Self-study group: Participants receive written materials and videos related to fire escape, and are allowed to search for relevant content through the internet.

- Controller group: Participants use the interaction method of controllers and undergo full process training in the VR system. Except for the interaction method, everything else is completely consistent with the MoCap group.

- MoCap group: Participants undergo training via the MIFE system, encompassing personalized knowledge assessments, complete scenario-based escape drills, and intelligent performance escape with real-time feedback.

Experimental stage: Each group was introduced to the experiment separately. To ensure that participants in the VR group could operate the VR system correctly, we provided them with system usage instructions. Notably, the instructions did not include any content related to fire escape drills.

Evaluation and post-test: After completing their respective training protocols, all participants were required to perform escape operations in a simulated real-world scenario. During the evaluation procedure, professional evaluators quantitatively assessed participants' performance. Subsequently, all participants completed a fire safety knowledge questionnaire B containing 10 multiple-choice questions. Additionally, structured interviews and system questionnaires are only administered to the MoCap group, because the self-study group does not experience the MIFE system, and the controller group use a different interaction method. The experimental process is shown in Fig. 5.

5.1.3 Evaluation Criteria and Metrics

Objective measurements: We use 3 indicators as evaluation criteria: knowledge mastery, posture score, and performance score. 1) Knowledge mastery: We use knowledge questionnaires during the evaluation stage to assess their mastery of fire safety theory knowledge. Compared with baseline results from questionnaire A obtained during the preparation stage, the improvement in theoretical knowledge for each participant is quantitatively calculated. 2) Posture score: Posture score is obtained by professionals based on participants' postures during the evaluation stage, with a maximum score of 10 points. 3) Performance score: Performance score is evaluated by professionals based on participants' reaction time, path selection, and facial expression during the evaluation stage.

For the above indicators, we conduct Shapiro-Wilk test ($\alpha = 0.05$) and Levene test ($\alpha = 0.05$) on each set of data to evaluate the assumptions of normality and homogeneity of variances, respectively. For data that satisfy both assumptions, we perform a one-way ANOVA ($\alpha = 0.05$) followed by pairwise comparisons using Tukey's Honestly Significant Difference (HSD) post hoc test ($\alpha = 0.05$). Otherwise, we employ the non-parametric Kruskal-Wallis test ($\alpha = 0.05$), followed by Dunn's post hoc test with Bonferroni correction ($\alpha' = 0.0167$) to adjust for multiple comparisons.

Subjective measurements: The post-test system questionnaire consists of 2 parts: usability and user experience.

Usability (The second version of Post-study System Usability Questionnaire [51], PSSUQ): We utilize the PSSUQ, which consists of 19 questions, to directly assess the usability of the MIFE system. This scale quantifies participants' experiences while interacting with the system, evaluating its usability and ease of use.

User Experience (User Experience Questionnaire-Short [52], UEQ-S): Participants complete the UEQ-S to assess the overall user experience. The UEQ-S scale consists of eight questions, divided into two dimensions: pragmatic quality and hedonic quality.

5.2 Result

This section presents the quantitative and qualitative experimental results from the user study. To ensure data integrity, we apply rigorous data cleaning procedures and the verification results show that all data is retained.

5.2.1 Objective Results

The descriptive statistics on knowledge mastery, posture score, and performance score are shown in Tab. 2, which includes mean, standard deviation, and median.

Knowledge mastery: The average scores for questionnaire A are 80, 68, and 67, accompanied by standard deviations of 12.47, 12.29, and 9.49, respectively, corresponding to the self-study group, controller group, and MoCap group. In the same order, the average scores are 84, 86, and 89, with standard deviations of 8.43, 13.50, and 7.38 for questionnaire B. As shown in Fig. 6 (a), the average score of questionnaire B in the controller group and the MoCap group is significantly higher than that of questionnaire A.

Further, the difference between the scores of questionnaire B and questionnaire A follows a normal distribution and homogeneity as shown in Tab. 2. Subsequently, a one-way ANOVA reveals significant differences in knowledge mastery across groups ($F_{2,27} = 5.853$, $p = 0.008$, $\eta^2 = 0.302$). HSD post hoc test shows that the score of the MoCap group is significantly higher than the self-study group ($p = 0.010$, Cohen's d: 1.310) and the controller group is significantly higher than the self-study group ($p = 0.034$, Cohen's d: 1.103). There is no significant difference between the MoCap group and the controller group ($p = 0.856$, Cohen's d: 0.230). As shown in Fig. 6 (b), the 95% simultaneous confidence intervals for each group are as follows: [-2.99, 10.99] for the self-study group; [12.01, 25.99] for the controller group; and [15.01, 28.99] for the MoCap group.

Posture score: The posture score follows a normal distribution and homogeneity as shown in Tab. 2. Subsequently, a one-way ANOVA reveals significant differences in posture score across groups ($F_{2,27} = 18.110$, $p < 0.001$, $\eta^2 = 0.573$). HSD post hoc test shows that the score of the MoCap group is significantly higher than those of the controller group ($p < 0.001$, Cohen's d: 2.134) and the self-study group ($p < 0.001$, Cohen's d: 2.265). There is no significant difference between the controller group and the self-study group ($p = 0.586$, Cohen's d: 0.419). As shown in Fig. 6 (c), the 95% simultaneous confidence intervals for each group are as follows: [4.55, 6.05] for the self-study group; [5.15, 6.65] for the controller group; and [7.95, 9.45] for the MoCap group.

Performance score: As shown in Tab. 2, the mean score of the MoCap group is 3.08 points higher than the self-study group and 2.38 points higher than the controller group. Meanwhile, the performance score does not follow a normal distribution. Subsequently, the Kruskal-Wallis test reveals a significant difference across groups ($H = 20.072$, $p < 0.001$, $\varepsilon^2 = 0.692$). Dunn's post hoc test with Bonferroni correction further identifies significant differences between groups (self-study group vs. controller group: $Z = -7.366$, $p < 0.001$; self-study group vs. MoCap group: $Z = -41.783$, $p < 0.001$; controller group vs. MoCap group: $Z = -34.417$, $p < 0.001$), as shown in Fig. 6 (d).

5.2.2 Subjective Results

Usability: Tab. 3 presents the descriptive statistics of participants' scores on metrics including system usefulness, information quality, interface quality, and overall usability. A comparison between the scores obtained by the system and the normative reference values

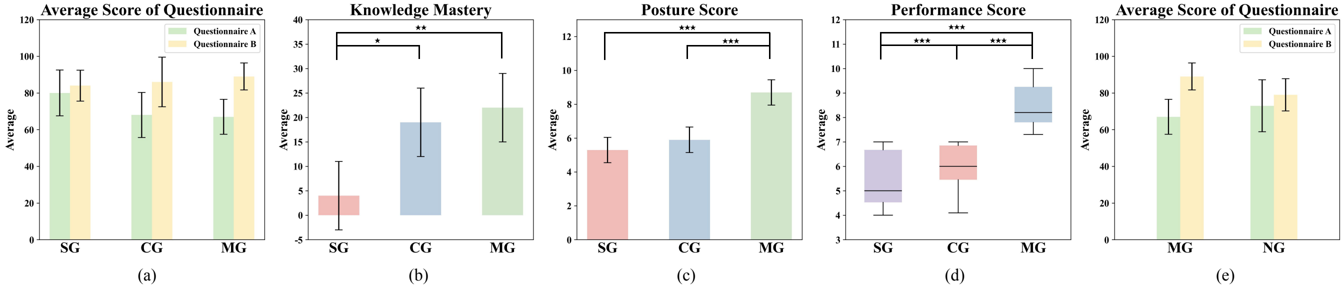


Figure 6: Objective results. (a) Average score of questionnaire A and questionnaire B with standard deviation. (b) 95% simultaneous confidence intervals of knowledge mastery based on Tukey HSD. (c) 95% simultaneous confidence intervals of posture score based on Tukey HSD. (d) Distribution of performance score and significance of Dunn's post hoc test with Bonferroni correction. (e) Average score results between the MoCap group and the non-KG group. (SG: Self-study Group, CG: Controller Group, MG: MoCap Group, NG: non-KG Group.)

Table 2: Statistical analysis of experimental data. (KM: Knowledge Mastery, PoS: Posture Score, PeS: Performance Score.)

Group	**Mean**			**Standard Deviation**			**Median**			**Shapiro-Wilk Test (p-value)**			**Levene's Test**		
	KM	PoS	PeS	KM	PoS	PeS	KM	PoS	PeS	KM	PoS	PeS	KM	PoS	PeS
Self-study Group	4	5.3	5.38	16.47	1.42	1.23	5	5.5	5.0	0.532	0.520	0.061	✔	✔	✔
Controller Group	19	5.9	6.08	**9.94**	1.45	1.05	**20.0**	6.0	6.35	0.152	0.119	**0.027**			
MoCap Group	**22**	**8.7**	**8.46**	10.33	**1.16**	**1.03**	**20.0**	**9.0**	**8.2**	0.191	0.124	0.123			

[51] for each metric reveals that MIFE significantly outperforms the reference scores, indicating its high level of usability.

Table 3: The score of PSSUQ from the MoCap group.

Observation Indicators	Reference	System Score	Result
Overall usability	3.02	5.41	high
System usefulness	3.02	5.48	high
Information quality	3.24	5.29	high
Interface quality	2.71	5.63	high

User experience: Tab. 4 presents the descriptive statistics of participants' scores on metrics including pragmatic quality, hedonic quality and overall. The results show that the MIFE system can provide a good user experience and participants' subjective feelings are overall good, but there are slightly more individual differences in hedonic quality than in pragmatic quality.

Table 4: The score of UEQ-S from the MoCap group.

Observation Indicators	Average	Std Dev	Min	Max
Pragmatic quality	2.53	0.28	0	3
Hedonic quality	2.50	0.41	1	3
Overall	2.51	0.25	0	3

5.2.3 Post-test Interview

At the end of the experiment, we conducted interviews with ten participants from the MoCap group to gather feedback on their experience using the MIFE system. Participants generally agreed that the interaction method of the MIFE system is highly novel, as it not only eliminates the reliance on traditional handheld controllers but also achieves high-precision full-body motion capture, resulting in more authentic and natural motion replication. Regarding immersion, they consistently affirmed that the MIFE system, through the combined effects of high-quality scene rendering and multimodal feedback mechanisms, creates a highly realistic and convincing virtual environment, enhancing their sense of presence and engagement during the experience. Furthermore, two participants specifically mentioned the system's posture error feedback function, noting that it effectively helps them promptly detect and correct movement deviations. In terms of knowledge recommendation powered by the large model, participants widely appreciated its ability to deliver relevant yet unfamiliar knowledge based on their operational data. However, four of them pointed out that the current system exhibits some latency in the large model's analysis response and expressed a hope for improved real-time performance and reduced waiting time in future versions. For effectiveness, two participants expressed concern that training through the VR system may make it difficult to evaluate long-term outcomes. In contrast, the others believed that the inherent advantage of the VR system lies in its reusability, allowing repeated training sessions to reinforce memory retention. Thus, they remained optimistic about long-term effectiveness and recommended further exploration of more efficient system functionalities.

In addition, We randomly selected three participants each from the other two groups to ask if they would be willing to try the MIFE system, and all six expressed strong interest. This shows that the MIFE system, with its novel interaction method and highly immersive experience, has broad appeal and is likely to attract more users.

6 Ablation

Ablation 1: Personalized recommendations of knowledge graph enable participants to master more fire safety knowledge.

To evaluate dynamic personalized recommendations powered by large models and knowledge graphs, we recruit ten new participants to form a non-KG group. The experimental procedure is identical to that of the MoCap group in the user study, with the sole exception that no dynamic knowledge recommendations are provided. The evaluation indicator adopts the difference between questionnaire B and questionnaire A.

In the non-KG group, the average score for questionnaire A is 73 (SD = 14.18), and for questionnaire B it is 79 (SD = 8.76). The comparison between the non-KG group and the MoCap group is shown in Fig. 6 (e). The non-KG group shows a 6-point difference between questionnaire B and questionnaire A. This result is similar to the 4-point difference observed in the self-study group, but is markedly lower than the 22-point gap seen in the MoCap group. These findings suggest that personalized recommendations supported by knowledge graphs effectively enhance participants' acquisition of unfamiliar knowledge.

Ablation 2: Temperature feedback and odor feedback can provide a more immersive experience.

To evaluate the effects of odor feedback and temperature feedback on user immersion, participants from the MoCap group were invited to repeat the training stage without multimodal feedback after completing the initial user study and were subsequently interviewed for further insights.

They reported perceiving changes in temperature and odor, particularly during fire spread. The noticeable rise in temperature and distinct irritating odor immediately evoked nervousness, which enhanced immersion and supported successful escape within the simulated fire environment. However, three participants noted that the pungent odor did not accurately resemble the smell of burning materials, a discrepancy that reduced the impact on their experience.

Ablation 3: Validate the effectiveness of the fire spread model.

For fire-spread simulation, we have validated the effectiveness of the fire spread model through expert review. We have invited experts from the Emergency Management Department, including representatives from the Tianjin Fire Research Institute, the Shanghai Fire Research Institute, and the Beijing Fire Rescue Corps, to evaluate the fire spread model. The experts unanimously agreed that the fire spread simulation realistically replicates the actual process of fire spread, thereby enhancing the immersive experience.

7 Discussion

7.1 Analysis of User Study Results

The results of the user study indicate that, compared to the self-study method and VR controller-based interactive systems, the MoCap group achieves superior outcomes in terms of knowledge mastery, posture score, and performance score. This advantage can be attributed to its dynamic personalized knowledge recommendation mechanism, which adapts to trainees with different knowledge backgrounds. By replacing conventional VR controller-based interaction, motion capture technology eliminates dependence on handheld devices and enables real-time mapping of the trainees' actual movements onto a virtual avatar. Furthermore, the immersive experience provided through an eight-mode omnidirectional environment allows trainees to engage in realistic fire scenarios, evoking an authentic sense of urgency and on-site responsiveness.

About user experience, the results indicate that MIFE achieves a balanced experience in user experience, and participants all believe that the product is very easy to use and reliable. It is not only efficient, clear, and reliable in completing tasks, but also attractive, novel, and exciting, which can bring a sense of pleasure to trainees. In terms of hedonic quality, although participants' subjective feelings are overall good, there are slightly more individual differences than in pragmatic quality. However, this is a normal finding as aesthetic and emotional preferences are more subjective.

7.2 The Necessity of an Extensively Equipped System

For the fire escape training system, it is necessary to construct a quasi-photorealistic, all-encompassing simulation. The fire escape skills involve a complex integration of multi-level knowledge, such as behavioral memory, situational judgment, and stress response, and their effective mastery highly depends on high simulation training conditions. Firstly, the internalization of behavioral knowledge depends on the authenticity of action mapping. When the virtual environment is highly consistent with real actions, trainees form a stable action memory through proprioceptive and visual feedback. Secondly, the construction of situational cognitive ability requires the support of multisensory collaborative stimulation and intelligent interaction. Through the comprehensive feedback of thermal, olfactory, and auditory channels, the system can simulate the dynamic environment of a real fire scene. At the same time, eye tracking can capture the visual attention path of trainees, facial expression recognition can monitor changes in their psychological state, and language interaction provides a natural communication channel for knowledge answering. The synergistic effect of these modalities jointly trains trainees' attention allocation, risk assessment, and emergency decision-making abilities in complex situations.

In the user study, the comprehensive performance score of the MoCap group reached 8.46 points, which was 3.08 and 2.38 points higher than that of the Self-study group and the Controller group, respectively. This result demonstrates that the multimodal high-simulation design adopted by MIFE is a key factor in improving training effectiveness and promoting skill transfer.

8 Limitations

During the interviews, several participants noted that the simulated pungent odor differs from the smell of an actual fire. To address this issue, we believe that simulating the odors produced by burning various materials and then releasing either individual or blended scents as needed is crucial for achieving a more realistic olfactory experience. Moreover, four participants also noted that the waiting time for dynamic personalized knowledge recommendations powered by large models is somewhat lengthy. In future work, optimizing the inference process could help improve the response speed.

9 Conclusion

In this paper, we propose a new immersive fire escape training system, MIFE, which creates a highly immersive virtual fire scene to enable trainees to effectively master and apply key escape skills. Specifically, we have designed an 8-modal fusion perception module to enhance the immersive experience of trainees from multiple perspectives. Moreover, we utilize a dual-perspective motion capture module to eliminate the dependence on controllers, allowing trainees to complete escape action training without restrictions. Additionally, the real-time spread of fire enhances the authenticity of the scene, and the dynamic recommendation based on the knowledge graph can be adapted to different trainees for personalized training. User studies confirm that the training outcomes achieved with MIFE are superior to those obtained through self-study and controller-based interaction. This advantage is quantified by a performance improvement of 3.08 and 2.38 points, respectively, on a 10-point evaluation scale. Ablation experiments validate the effectiveness of multimodal feedback, dual-perspective motion capture, and knowledge graph components. These results enhance the potential for innovative fire evacuation drills and support the advancement of similar VR-based skill training systems.

Supplemental Materials

All supplemental materials are available on OSF at https://osf.io/aswqj/overview. In particular, they include (1) A PDF file contains the implementation details including heat modality, olfactory modality, full-body motion capture, hand tracking, fire spread model, and knowledge Recommendation, (2) Two PDF files contain questionnaire A and questionnaire B which are used in user study.

Acknowledgments

This work was supported in part by National Key R&D Program of China (2023YFC3082100), Natural Science Foundation of Tianjin (No.24JCQNJC01620), Science Fund for Distinguished Young Scholars of Tianjin (No.22JCJQJC00040).

References

[1] Center for Fire Statistics of CTIF. World fire statistics. Technical Report 30, International Association of Fire and Rescue Services, 2025. 1

[2] BJ Meacham. Global plan for a decade of action for fire safety. *International Fire Safety*, 2021. 1

[3] Audrie A Chavez, Sarah V Duzinski, Tareka C Wheeler, and Karla A Lawson. Teaching safety at a summer camp: Evaluation of a fire safety curriculum in an urban community setting. *Burns*, 40(6):1172–1178, 2014. 1, 2

[4] Paul H Lee, Baoguo Fu, Wangting Cai, Jingya Chen, Zhenfei Yuan, Lifen Zhang, and Xiuhong Ying. The effectiveness of an on-line training program for improving knowledge of fire prevention and evacuation of healthcare workers: A randomized controlled trial. *PLOS ONE*, 13(7):e0199747, 2018. 1, 2

[5] Hosan Kang, Jinseong Yang, Beom-Seok Ko, Bo-Seong Kim, Oh-Young Song, and Soo-Mi Choi. Integrated augmented and virtual reality technologies for realistic fire drill training. *IEEE Computer Graphics and Applications*, 44(2):89–99, 2023. 2, 3

[6] Muhammad Hasham Qazi, Farhan Khan, Jeeeun Kim, and Edgar J. Rojas-Munoz. Developing a VR-based training platform for emergency fire handling services using unity 3D. In *International Conference on Frontiers of Information Technology*, pages 102–107, 2023. 2, 3

[7] Jiaxin Ling, Xiaojun Li, Yi Shen, Chao Chen, Zhiguo Yan, Hehua Zhu, and Haijiang Li. Human centric VR system development supporting fire emergency evacuation: A novel knowledge-data dual driven approach. *Expert Systems with Applications*, 273:126895, 2025. 2, 3, 4

[8] Joana Oliveira, Joana Aires Dias, Rita Correia, Raquel Pinheiro, Vítor Reis, Daniela Sousa, Daniel Agostinho, Marco Simões, Miguel Castelo-Branco, et al. Exploring immersive multimodal virtual reality training, affective states, and ecological validity in healthy firefighters: Quasi-experimental study. *JMIR Serious Games*, 12(1):e53683, 2024. 2, 3

[9] Miguel Melo, Guilherme Gonçalves, Pedro Monteiro, Hugo Coelho, José Vasconcelos-Raposo, and Maximino Bessa. Do multisensory stimuli benefit the virtual reality experience? A systematic review. *IEEE Transactions on Visualization and Computer Graphics*, 28(2):1428–1442, 2022. 2

[10] Alison Crosby, MJ Johns, Katherine Isbister, and Sri Kurniawan. Designing FEVR: A VR game for wildfire evacuation readiness. In *Proceedings of the 20th International Conference on the Foundations of Digital Games*, pages 1–4, 2025. 2, 3

[11] Musaab H Hamed-Ahmed, Diego Ramil-López, Paula Fraga-Lamas, and Tiago M Fernández-Caramés. Towards an emotion-aware metaverse: A human-centric shipboard fire drill simulator. *Technologies*, 13(6):253, 2025. 2, 3

[12] Jo Skjermo, Claudia Moscoso, Daniel Nilsson, Håkan Frantzich, Åsa S Hoem, Petter Arnesen, and Gunnar D Jenssen. Analysis of visual and acoustic measures for self-evacuations in road tunnels using virtual reality. *Fire Safety Journal*, 148:104224, 2024. 2

[13] Endel Tulving and Donald M. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80:352–373, 1973. 2

[14] D., R., GODDEN, A., D., and BADDELEY. Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 1975. 2

[15] Muhammad Hafeez Basri. *Encoding specificity principle (ESP) in enhancing motor memory performance skills*. PhD thesis, Universiti Teknologi MARA, 2011. 2

[16] Xiaochun Zhang, Linjie Chen, Junhao Jiang, Yixin Ji, Shuyang Han, Ting Zhu, Wenbin Xu, and Fei Tang. Risk analysis of people evacuation and its path optimization during tunnel fires using virtual reality experiments. *Tunnelling and Underground Space Technology*, 137:105133, 2023. 2

[17] Stylianos Mystakidis, Jeries Besharat, George Papantzikos, Athanasios Christopoulos, Chrysostomos Stylios, Spiros Agorgianitis, and Dimitrios Tselentis. Design, development, and evaluation of a virtual reality serious game for school fire preparedness training. *Education Sciences*, 12(4):281, 2022. 2

[18] Andreas Marougkas, Christos Troussas, Akrivi Krouska, and Cleo Sgouropoulou. How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade. *Multimedia Tools and Applications*, 83(6):18185–18233, February 2024. 2

[19] Camelia Delcea and Liviu-Adrian Cotfas. Increasing awareness in classroom evacuation situations using agent-based modeling. *Physica A: Statistical Mechanics and Its Applications*, 523:1400–1418, 2019. 2

[20] Hana Najmanová, Petr Novák, and Enrico Ronchi. The status quo of fire evacuation drills in nursery schools. *Safety Science*, 191:106915, 2025. 2

[21] Zhian Huang, Rongxia Yu, Yang Huang, Jinyang Li, Hao Ding, Yukun Lei, Pengfei Wang, and Danish Jameel. Reliability analysis of a building real fire simulation training system. *Fire*, 6(10):369, 2023. 2

[22] Linjing Sun, Boon Giin Lee, and Wan-Young Chung. Enhancing fire safety education through immersive virtual reality training with serious gaming and haptic feedback. *International Journal of Human–Computer Interaction*, 41(9):5607–5622, 2025. 2

[23] Zelin Jiang, Shuhao Zhang, Yue Li, Ka Lok Man, Yong Yue, and Jeremy Smith. Is VR always a better choice? Investigating the effects of game modes and role-playing on fire escape simulation training. In *International Conference on Virtual Reality*, pages 338–346. IEEE, 2024. 2

[24] Yaqin Fu and Qi Li. A virtual reality–based serious game for fire safety behavioral skills training. *International Journal of Human–Computer Interaction*, 40(19):5980–5996, 2024. 2

[25] Nithya Shree T. and A. Grace Selvarani. Virtual reality based system for training and monitoring fire safety awareness for children with autism spectrum disorder. In *International Conference on Devices, Circuits and Systems*, pages 26–29, 2020. 2

[26] Shih-Yeh Chen and Wei-Che Chien. Immersive virtual reality serious games with dl-assisted learning in high-rise fire evacuation on fire safety training and research. *Frontiers in psychology*, 13:786314, 2022. 2

[27] Musaab H Hamed-Ahmed, Paula Fraga-Lamas, and Tiago M Fernández-Caramés. Towards the industrial metaverse: a game-based VR application for fire drill and evacuation training for ships and shipbuilding. In *International ACM Conference on 3D Web Technology*, pages 1–6, 2024. 2

[28] Philipp Braun, Michaela Grafelmann, Felix Gill, Hauke Stolz, Johannes Hinckeldeyn, and Ann-Kathrin Lange. Virtual reality for immersive multi-user firefighter-training scenarios. *Virtual reality & Intelligent Hardware*, 4(5):406–417, 2022. 2

[29] Ungyeon Yang, Hyungki Son, and Kyungsik Han. Developing a realistic VR interface to recreate a full-body immersive fire scene experience. In *SIGGRAPH Asia Posters*, pages 1–2. 2023. 2, 3

[30] David Narciso, Maximino Bessa, Miguel Melo, and José Vasconcelos-Raposo. Virtual reality for training-the impact of smell on presence, cybersickness, fatigue, stress and knowledge transfer. In *International Conference on Graphics and Interaction*, pages 115–121. IEEE, 2019. 2

[31] Zhang Bo, Sun Jun, Shang Lei, Hongji Xu, and Yang Cheng. Design and realization of ship virtual fire training system based on hmd. *Journal of System Simulation*, 31(1):43–52, 2019. 2

[32] Daniel Martin, Sandra Malpica, Diego Gutierrez, Belen Masia, and Ana Serrano. Multimodality in VR: A survey. *ACM Computing Surveys*, 54(10s), September 2022. 3

[33] Chuanzhi Su, Mengjie Huang, Jingjing Zhang, and Rui Yang. The application of eye tracking on user experience in virtual reality. In *International Conference on Cognitive Aspects of Virtual Reality*, pages 57–62, 2023. 3

[34] Baidu AI Cloud. Baidu speech recognition developer guide. https://cloud.baidu.com/doc/SPEECH/s/zk4o0bmzk, 2023. online document. 3

[35] International Organization for Standardization. Building construction—accessibility and usability of the built environment. ISO 21542:2021, 2011. 3

[36] Wout van Bommel. Emergency lighting. In *Encyclopedia of Color Science and Technology*, pages 787–791. Springer, 2023. 3

[37] International Organization for Standardization. Graphical symbols-safety signs-safety way guidance systems (iso 16069: 2004), 2004. 3

[38] Gianni Vercelli, Saverio Iacono, Luca Martini, Michele Zardetto, and Daniele Zolezzi. From risk to readiness: VR-based safety training for industrial hazards, 12 2024. 3

[39] Guido Makransky, Thomas S. Terkildsen, and Richard E. Mayer. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60:225–236, 2019. 3

[40] Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Yao Feng, and Michael J Black. Tokenhmr: Advancing human mesh recovery with a tokenized pose representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1323–1333, 2024. 4

[41] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, and Jitendra Malik. Reconstructing hands in 3D with transformers. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2024. 4

[42] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015. 4

[43] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics*, 36(6):1–17, November 2017. 4

[44] M. Byari, A. Bernoussi, O. Jellouli, M. Ouardouz, and M. Amharref. Multi-scale 3D cellular automata modeling: Application to wildland fire spread. *Chaos, Solitons & Fractals*, 164:112653, 2022. 5

[45] Yunyi Shen, Kaixiang Yi, Wenju Zhou, Minrui Fei, and Zehao Lv. The bert-bilstm-crf model applied to chinese entity recognition for the science and technology service field. In *Chinese Control Conference*, pages 4205–4210, 2022. 5

[46] M.A. Rodriguez and M.J. Egenhofer. Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15(2):442–456, 2003. 5

[47] Jinxiu Chen, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. Unsupervised feature selection for relation extraction. In *International Joint Conference on Natural Language Processing*, 2005. 6

[48] Hasan Abu-Rasheed, Christian Weber, and Madjid Fathi. Knowledge graphs as context sources for llm-based explanations of learning recommendations. In *IEEE Global Engineering Education Conference*, pages 1–5, 2024. 6

[49] Manuel Kaufmann, Jie Song, Chen Guo, Kaiyue Shen, Tianjian Jiang, Chengcheng Tang, Juan José Zárate, and Otmar Hilliges. EMDB: The electromagnetic database of global 3D human pose and shape in the wild. In *International Conference on Computer Vision*, 2023. 6

[50] Timo von Marcard, Roberto Henschel, Michael Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3D human pose in the wild using imus and a moving camera. In *European Conference on Computer Vision*, sep 2018. 6

[51] James Lewis. Psychometric evaluation of the pssuq using data from five years of usability studies. *International Journal of Human–Computer Interaction*, 14:463–488, 09 2002. 7, 8

[52] Martin Schrepp, Andreas Hinderks, and Jorg Thomaschewski. Design and evaluation of a short version of the user experience questionnaire (ueq-s). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6):págs. 103–108, 2017. 7