VAction: A Lightweight and Integrated VR Training System for Authentic Film-Shooting Experience

Shaocong Wang
Department of Computer Science and Technology
Tsinghua University
Beijing, China
wangsc23@mails.tsinghua.edu.cn

Che Qu
Film & TV Technology Department
Beijing Film Academy
Beijing, Beijing, China
quche@bfa.edu.cn

Minjing Yu*
College of Intelligence and Computing
Tianjin University
Tianjin, China
minjingyu@tju.edu.cn

Chao Zhou
Human Computer Interaction Technology and Intelligent Information Processing Laboratory
Institute of Software Chinese Acadamy of Sciences
Beijing, China
zhouchao@iscas.ac.cn

Yuntao Wang
Department of Computer Science and Technology
Tsinghua University
Beijing, China
yuntaowang@tsinghua.edu.cn

Yu-Hui Wen
School of Computer Science and Technology
Beijing Jiaotong University
Beijing, China
yhwen1@bjtu.edu.cn

Yuanchun Shi
Department of Computer science and Technology
Tsinghua University
Beijing, China
shiyc@tsinghua.edu.cn

Yong-Jin Liu*
Computer Science and Technology
Tsinghua University
Beijing, Beijing, China
liuyongjin@tsinghua.edu.cn

Abstract

The film industry exerts significant economic and cultural influence, and its rapid development is contingent upon the expertise of industry professionals, underscoring the critical importance of film-shooting education. However, this process typically necessitates multiple practice in complex professional venues using expensive equipment, presenting a significant obstacle for ordinary learners who struggle to access such training environments. Despite VR technology has already shown its potential in education, existing research has not addressed the crucial learning component of replicating the shooting process. Moreover, the limited functionality of traditional controllers hinder the fulfillment of the educational requirements. Therefore, we developed *VAction* VR system, combining high-fidelity virtual environments with a custom-designed controller to simulate the real-world camera operation experience. The system's lightweight design ensures cost-effective and efficient deployment. Experiment results demonstrated that *VAction* significantly outperforms traditional methods in both practice effectiveness and user experience, indicating its potential and usefulness in film-shooting education.

*Minjing Yu and Yong-Jin Liu are corresponding authors.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3714217

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; • **Applied computing → Interactive learning environments**.

Keywords

Film Production Education, Virtual Reality, Training System, Interaction Design

ACM Reference Format:
Shaocong Wang, Che Qu, Minjing Yu, Chao Zhou, Yuntao Wang, Yu-Hui Wen, Yuanchun Shi, and Yong-Jin Liu. 2025. VAction: A Lightweight and Integrated VR Training System for Authentic Film-Shooting Experience. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3706598.3714217

1 Introduction

The film industry, with its profound economic and cultural impact globally, necessitates the cultivation of skilled cinematographers for its sustained and positive development [6]. However, the development of professional film-shooting skills requires extensive practical training [17]. This practice typically demands access to specific environments and specialized equipment, which significantly increases costs. For example, the average price of professional filming equipment can exceed several thousand dollars, while studio rental fees can range from hundreds to thousands of dollars per day. These costs, coupled with complex setup procedures and the need for iterative practice, create significant barriers for students in traditional educational settings. Similar challenges are seen in other skill-based training, such as surgical procedures, where realistic practice is vital.

To address these challenges, there has been growing interest in utilizing Virtual Reality (VR) technology to simulate real-world environments, offering a cost-effective and flexible alternative for practice [15, 24, 25, 48]. VR has already proven to be effective in fields like medical education, where it provides learners with the opportunity to engage in simulated practice without the need for extensive physical resources, thus helping to mitigate the steep learning curve associated with these skills [14, 46, 61].

Although VR technology has been widely applied in film production, especially in previsualization [40–42, 55], its application in film-shooting education is still limited. Most existing VR research in film production education has focused on improving visual immersion and overall user experience [29, 39, 60], rather than replicating the crucial aspects of the shooting workflow and the authentic feel of operating a real camera. Furthermore, current VR systems typically rely on gestures or standard controllers, which fail to encompass the full functionality of a camera and whose operational logic does not align with that of a real camera. This discrepancy makes it challenging to accurately replicate the real-world experience of using a camera. To address these gaps, we have developed a system that integrates both software and hardware, aiming to provide learners with a more authentic and practical experience.

In this paper, we developed an effective VR film-shooting training system called *VAction*, which integrates both software and hardware components. Based on surveys and interviews with film-shooting students and teachers, this study has identified key design principles focusing on high-fidelity visual effects, preservation of operational habits, and lightweight design.

On the hardware side, we immerse users in a virtual environment using a standard commercial VR headset. To replicate the tactile feel and operational characteristics of a professional camera, we developed a custom controller specifically designed for camera operators. This controller ensures that physical interactions in the virtual environment closely resemble those in the real world, thereby preserving operational habits that are crucial for effective skill training. The software component of this VR system is built on a lightweight architecture that allows for rapid deployment and easy access, ensuring efficiency and adaptability to various training needs. We utilized computer vision and laser scanning technologies to reconstruct the training environment, quickly providing a high-fidelity virtual environment that closely mirrors real-world conditions. The complex functionality of professional cameras has been modularized, allowing users to focus on targeted training based on their specific needs, thereby reducing the learning curve.

We conducted a user experiment to validate the effectiveness of our VR film-shooting training system. A total of 24 beginners were recruited and divided into two groups: one using traditional learning methods and the other using our VR system for learning basic camera operation skills. The analysis of the operational data collected from both groups revealed that *VAction* significantly outperformed traditional methods in both learning outcomes and user experience, demonstrating the system's effectiveness. Additionally, our system has been implemented in a local television station, and the practice showed that the efficiency of live broadcast previsualization has been greatly improved.

Specifically, this paper presents the following contributions:

- Based on interviews with professionals, we identified the challenges faced by film production education, including limitations in specialized equipment and venues, as well as insufficient practical opportunities. We proposed basic design principles that should be followed when utilizing VR technology to address these challenges: high-fidelity visuals, preservation of operational habits, and lightweight design.
- We developed *VAction*, a VR film-shooting training system that combines high-fidelity virtual scenes and shots with a self-developed controller, accurately replicating the operational processes and tactile feel of a real camera. The system is built on lightweight principles, allowing for quick deployment and convenient access to various modules and functional expansions.
- Through a user experiment, we demonstrated the effectiveness of *VAction* in terms of learning outcomes and user experience. By providing an immersive learning environment and embodied operational experiences, it promotes better transfer of learning.

2 Background and Related Works

In this section, we present the related work on virtual reality technology for learning and training, traditional and advanced methods in film production educaiton, and transfer learning and embodied cognition theory.

2.1 Virtual Reality Technology for Learning and Training

Digital technology, especially VR technology, has gradually become a key component in the field of education. VR is a technology that transports users to a virtual environment through specialized virtual headphones and visual devices, providing them with an immersive experience [1, 16, 35, 48].

The immersion and interactivity offered by VR technology significantly enhance the learning experience, providing students with unique perspectives [47]. In recent years, numerous educational methods and learning theories have been applied to VR: constructivist learning [18, 20] highlights active knowledge construction through interaction; experiential learning [28] emphasizes learning through concrete experiences; gamification [2] enhances engagement and motivation; John Dewey's learning-by-doing theory [10] underscores the importance of practical engagement; and flow theory [2, 9] suggests that immersive environments can facilitate deep focus and enjoyment. These theories all emphasize a learner-centered approach, where learners acquire knowledge and skills through practice and experience in real or simulated contexts [37, 45, 49].

VR's immersive environments, sensory enhancement, and interactive engagement support experiential learning that enhances students' motivation, participation, and comprehension, while providing educators with innovative teaching tools that increase flexibility and efficiency in instruction [24, 25]. In addition, VR provides an unconstrained virtual work environment, avoiding the inherent limitations and distractions of traditional work settings [15]. This makes VR a powerful tool for training complex skills that require specific environments and equipment [16], such as surgical procedures and operations in high-risk industrial settings. In surgical education, VR enables safe and immersive simulations that enhance the anatomical understanding of the interns and their surgical confidence [46, 61], and can even be integrated with Augmented Reality (AR) for remote guidance and collaboration [14]. Similarly, VR technology is increasingly used in trainings that involve hazardous environments such as fire safety [52], industrial training [3], and aircraft piloting [44]. The evolution of VR from an entertainment medium to a critical training technology underscores its effectiveness in providing realistic, risk-free practice environments across various fields.

It is worth noting that the learning-by-doing theory faces certain limitations in VR due to the lack of physical interaction [37]. To fully harness the potential of Dewey's theory [10] in VR, researchers and practitioners must find creative solutions to overcome these limitations, such as incorporating multi-dimensional feedback or integrating both software and hardware elements.

2.2 Traditional Approaches and Advanced Technologies in Film Production Education

Traditional film production education combines theoretical learning with practical exercises [6]. Students learn camera operation, lighting, and editing through lectures and apply these skills in real-world settings. However, this approach is limited by high costs, restricted access to equipment, time restrictions, and limited practice opportunities.

As the integration of film and television production with digital technology becomes increasingly tight [38, 41, 42, 55, 60], artificial intelligence (AI) and computer graphics technologies have been introduced to address these challenges. For instance, AI visualization is used for better information delivery [12, 19, 58], 3D digital reconstruction can accurately replicate scenes and characters from films [39] to build VR environments, which further provide more accessible and cost-effective platforms for tasks like pre-visualization and virtual set design. Machine learning methods can be used to analyze and replicate real film camera movements to intelligently generate virtual camera motions, enhancing automation and stylization in 3D animation and virtual cinematography [22, 23]. These technologies allow for extensive practice without the need for expensive physical resources, making it easier for non-professionals and starters to get started with practice [40]. Additionally, they enable remote collaboration, enhancing the flexibility and reach of film production education [13, 29].

Although virtual environments hold significant potential in film education, current research has focused primarily on technological development and improving user experience, with a lack of in-depth comparative studies with traditional teaching methods [13]. Future research should be guided by educational objectives, integrating virtual technologies more effectively into teaching processes. For example, while these technologies offer many conveniences, they have not yet fully addressed the demands for operating real equipment in film production education. Numerous studies have highlighted that authentic tactile feedback is crucial for students to effectively master camera work [17, 65]. However, there is still no training system that fully meets the requirements for providing this essential hands-on experience.

In this paper, we present *VAction*, a VR film-shooting training system that faithfully reproduces the film production workflow by combining high-fidelity scenes with hardware controllers that simulate the tactile experience of real operations.

2.3 Transfer Learning Theory and Embodied Cognition Theory

In film production education, there is a need to cover conceptual knowledge such as camera structure and shooting technique theories, as well as procedural knowledge related to camera operations [5]. According to ACT theory [4], skill learning is divided into two stages. First, rules enter the learner's propositional network in the form of declarative knowledge, which includes facts and conceptual knowledge. Then, through practice, this knowledge is transferred to procedural knowledge. This process emphasizes the migration of theoretical knowledge to operational skills and highlights the critical role of practice in this knowledge transfer.

At the same time, the theory of transfer of learning — defined as "prior learning affecting new learning or performance" — is a core educational concept [36, 51]. Effective learning transfer, particularly in the early stages of skill acquisition, relies on bridging abstract theoretical knowledge with concrete operational tasks. This bridging process often requires contextualized learning and systematic practice, and the effectiveness of learning transfer is significantly influenced by the similarity between learning environments [11, 36]. Specifically, in film production education, the more similar the learning environments are (such as the arrangement of teaching spaces and personnel), the more effectively students can utilize relevant cues for learning or problem-solving. However, due to the challenge of consistently providing real-world scenes and cameras for training, students face difficulties in transferring the conceptual knowledge learned in the classroom to specific procedural operational knowledge.

Immersive VR technology in film production education can enhance learning outcomes and transfer abilities by simulating real-life scenarios. Studies have shown that the immersive experience provided by VR enables students to better grasp abstract concepts that are difficult to convey through traditional teaching methods [59]. The psychological and emotional congruence of the VR environment is key to successful transfer [50], echoing the positive impact of learning context and rule similarity on transfer effects emphasized in learning transfer theories. Developing such environments maximizes the potential of VR-based training, helping students to transfer theoretical knowledge more effectively to practical operations and thereby improving their mastery of procedural knowledge.

Embodied cognition theory suggests that there is a connection between motor and visual processes; the clearer the connection, the better the learning outcomes. The process of bodily movement can facilitate the processing and understanding of abstract concepts [30], especially when the physical interaction aligns with the visual characteristics of the concepts [21], such as manipulating objects for rotation and movement. Research indicates that

Table 1: Results for the challenges in traditional teaching mode

Theme	Sub-theme	Example
Challenges in Traditional Film-shooting Education	Disconnection between Theoretical Knowledge Acquisition and Practical Skill Development	Learning to operate the machine through videos and illustrations is simple, but mastering it requires extensive practice (P2); Traditional video learning is too monotonous. It would be ideal if, after each short segment, I could immediately practice hands-on, like in a tutorial for beginners (P3); Teachers spend multiple lessons on explaining each module of the equipment, followed by a single practice session, which is insufficient for mastering complex devices; it would be better to practice each module individually right after each lesson (P4).
	Constraints on Equipment Access and Practical Training Opportunities	Few courses focus on camera skills practice; instead, students mostly gain experience through short film group projects, where only one can handle the camera, limiting practice for all (P1); I mostly practice with a smartphone or SLR camera, but their operations are very different from professional equipment, which is hard to borrow due to high demand (P2). Access to the latest gear typically requires a well-funded team or commercial production, making it difficult for students in regular courses. While students can borrow equipment from the school, it's limited and in high demand (P4).
	Perceived Complexity and Initial Overwhelm with Professional Equipment	Professional equipment, with its many features and complex specifications, can be intricate to set up. Students, while eventually becoming familiar with it, often initially feel overwhelmed and find it challenging to operate (P1); Professional equipment is much heavier and larger than everyday devices, making me hesitant to operate it freely (P2); Professional equipment isn't overly complex, but it intimidates beginners (P4).

embodied learning experiences can enhance the transfer of learning, particularly for procedural knowledge [34], and VR is a good medium to provide such embodied experiences [27]. In addition to the experiences brought about by bodily movements in VR, the controller design must also induce embodiment through meaningful and consistent movements related to the content being learned, thus enhancing the transfer effect [26]. For example, haptic controllers leveraging tactile and force feedback facilitate embodied experiences, enhancing immersion and interaction in VR [53], while haptic canes offering multimodal feedback improve the immersive experience for navigation in virtual environments [64].

Through a literature review, we found that the application of VR in current film production education is limited, with a primary focus on the presentation of scenes and visuals [13]. There is a lack of attention to the training of film-shooting processes and skills, which are areas where traditional educational methods fall short. We aim to develop a VR teaching and training system specifically for the film-shooting training segment of film production education to help address the challenges currently faced in the field.

3 Formative Study

Our system is primarily designed to assist beginners in acquiring film-shooting skills. Before constructing the system, we conducted semi-structured one-on-one interviews with a film production teacher (P1) with over eight years of teaching experience, two beginners (P2, P3) who have just begun their study, and a film technology graduate student (P4) with five years of professional study experience. The reason for selecting such interviewees is that we intended not only to understand the difficulties of beginners but also to gain more comprehensive and enriched insights from more experienced educators and learners. This approach allows our system to align with the overall process and requirements of film shooting while providing a better experience for beginners in the details. Two of the interviews were conducted face-to-face, while the other two were held via online meetings, depending on the interviewees' preferences. Each interview lasted between 30 and 40 minutes. The discussions primarily focused on the key aspects and challenges in the current learning and training process of film-shooting techniques, as well as the critical features that a VR film-shooting training system should incorporate. All interviews were conducted in Chinese and recorded with each participant's consent. Upon completion of the interviews, each participant received $5 compensation.

3.1 Interview Results and Findings

During the interview process, we first inquired about the traditional film-shooting teaching mode and the experience of using equipment. From the interview results, we found that the high cost of professional equipment limits the practical aspects of traditional teaching, and students often took only one practical class after several theoretical classes (P4). Most of the students' hands-on experience with cameras comes from collaborative short film projects, with little time to use professional equipment individually (P1, P2). This leads to a gap between theoretical knowledge and practical skills (P2). The limited access to professional equipment and its high cost create psychological pressure for students, especially novices (P1, P4).

We then guided interviewees to envision technologies and methods to resolve current challenges. Three (P1, P2, P3) immediately thought of VR, while another (P4) mentioned digital scene applications. They all agreed that virtual scenes and simulated cameras could reduce dependency on physical setups and that VR could

Table 2: Results for supporting film-shooting education with VR

Theme	Sub-theme	Example
Feasibility of Supporting Film-shooting Education with VR Technology	Real-time Feedback and Efficient Learning Integration through VR	VR can provide updated feedback methods, such as showing relevant knowledge during practice and offering performance feedback afterward (P1); Unlike videos, which lack immediate responses, VR enables the system to directly manage interactions, offering real-time supervision and feedback that traditional methods cannot. VR can enhance involvement and allows me to verify the knowledge I just learned (P3).
	Convenient Access to Professional Equipment and Environments via VR	VR can offer a variety of scenes and spaces to choose from, enabling us to create many case studies with different set designs (P1); It allows for easy camera switching by adjusting parameters, unlike real-life camera setups that require purchasing different lenses (P2); If a school lacks enough professional cameras or a studio, VR can simulate scenes and cameras. Even if the effect is 80% of the real thing, it's still excellent, especially for beginners (P4).
	Potential Disadvantages or Challenges of VR	Motion sickness and the discrepancy between VR and real-world scenarios could impair educational outcomes (P1); The difference between VR controllers and real equipment might lead to incorrect muscle memory (P2, P3); Discrepancies between virtual and real feedback could be problematic. For professionals, even slight differences in visual aspects can matter (P4).
	Key Features That A VR-based Training System Should Achieve	The system should align with the real shooting process, offer a purely virtual environment without distractions, provide a variety of scenes, and ensure good interaction (P1); It's crucial to maintain focus, ensure effective knowledge transfer, and allow extensive practice at low cost (P3); The system must be accurate, simulate the real shooting process, and offer a similar feel to real equipment. The highest requirement is visual consistency with the real world (P4).

provide a more immersive experience. However, the interviewees (P2, P3, P4) also raised concerns about the differences between practicing with VR and real cameras, offering valuable insight for system development. Finally, we asked them to outline key features for such a digital or VR training system to guide our design.

We conducted thematic analysis on the interview transcripts [8]. All interview recordings provided in supplementary materials were transcribed into text using WeMeet, and then translated into English by GPT-4, followed by a double check of a English native speaker who is also familiar with Chinese well. A researcher manually segmented the materials based on the themes discussed. Finally, two researchers independently coded the interview materials, derived several themes through inductive analysis, and then merged and adjusted these themes through discussion. The final thematic analysis results are presented in Table 1 and Table 2.

3.2 Summary of Design Requirements

Based on the interview results, we propose the following primary design principles for the system: high-fidelity visuals, preservation of operational habits, and lightweight design, all of which are closely related to the overall theme of the system design.

High-fidelity visuals

A key aspect of the system design theme is to provide an effective practice environment. One of the advantages of a VR system is the ability to switch scenes and camera parameters conveniently, allowing for repeated low-cost practice. In line with this, both the teacher (P1) and the experienced student (P4) emphasized that to ensure the effectiveness of practice, the consistency of scenes and visuals with reality must be maintained. This is essential to avoid a disconnect in the learning process and aligns with the overall goal of creating a realistic and useful practice platform.

Preservation of operational habits

The overall design theme of our system aims to enhance real - world camera - using skills through virtual practice. Although students can engage in extensive practice within VR, the ultimate goal is to improve their performance with real cameras. Therefore, in keeping with this theme, it is crucial to ensure that the operational processes and tactile feedback in VR closely match those in real-world filming. This principle, highlighted by P2 and P3, is particularly important for beginners and is in line with the system's objective of effectively transferring skills from the virtual to the real environment.

Lightweight design

The system design theme includes leveraging the benefits of a virtual system to improve the training experience. We can implement functions that are difficult to achieve in real-world training, providing greater convenience for both teachers and students. In particular, through lightweight design, we can reduce the time and financial costs associated with film-shooting practice. For example, the rapid creation and switching of scenes and lenses enables teachers to quickly set up and deploy a training scenario. Additionally, by modularizing the various components of the camera, students can focus on specific modules according to the course content, thereby avoiding the psychological pressure of dealing with large and complex equipment (P4), and this contributes to the overall goal of creating an accessible and effective training system.

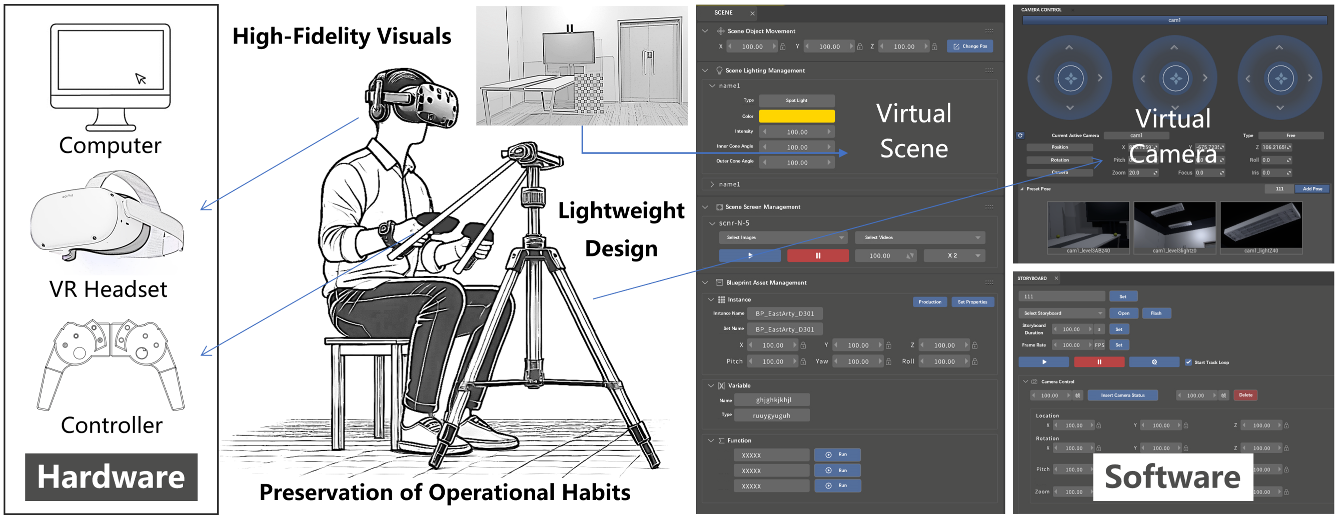


Figure 1: The VR film-shooting training system, *VAction*, consists of both hardware and software components. On the left side are the main hardware components, and the entire system operates on a single computer. Users wear a VR headset to enter the virtual scene and control the virtual camera using a controller. On the right side are the main functional modules of the system within the VR environment, including scene control, camera control, and camera movement modules.

4 System Design

In the previous section, we summarized the insights gathered from interviews, leading to the identification of three critical principles essential to the system's design: high-fidelity visuals, preservation of operational habits, and lightweight design. Building on these fundamental principles, we built *VAction*, a VR film-shooting training system. In this section, we dive into the specific design strategies and technological approaches implemented to ensure that each of these key aspects is effectively realized within the system.

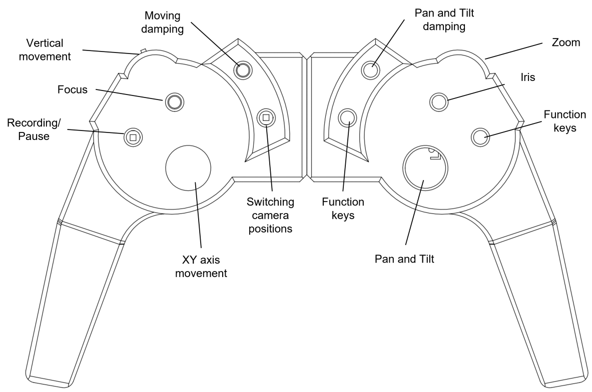
Unlike traditional learning modes which often separates video study from hands-on camera practice, users are immersed in *VAction*'s VR environments, practicing camera skills with controller handles that simulate real-hand feel. *VAction* liberates users from the constraints of location and expensive equipment. Furthermore, the modular design of the camera functions allows for convenient targeted training, reducing the psychological stress associated with complex equipment.

The VR film-shooting training system is composed of both hardware and software components as shown in Fig. 1. The hardware includes a computer to run the system, a video capture card to calibrate parameters of real camera lens (by streaming footage to the computer), a pair of hardware controllers for physical interaction, and a VR headset for display. The hardware part mainly corresponds to the display of images and the operation of the camera, including Zoom and Focus adjustment, and pan-and-tilt movement. The functional modules of the software include high-fidelity scene reconstruction and lens calibration, environment management, virtual camera control, shot and camera movement recording, storyboard generation, and digital asset management (covering models, lenses, and other camera components). The software part mainly corresponds to the various processes of operating the virtual camera for shooting, including scene and lens preparation, camera type selection, camera operation for shooting and recording. We have also implemented many commonly used auxiliary components of professional cameras in the form of Blueprint functions. All of these functions are accessible through an interface built on Unreal Engine 4.27, which is displayed simultaneously with the virtual camera's image in the VR scene.

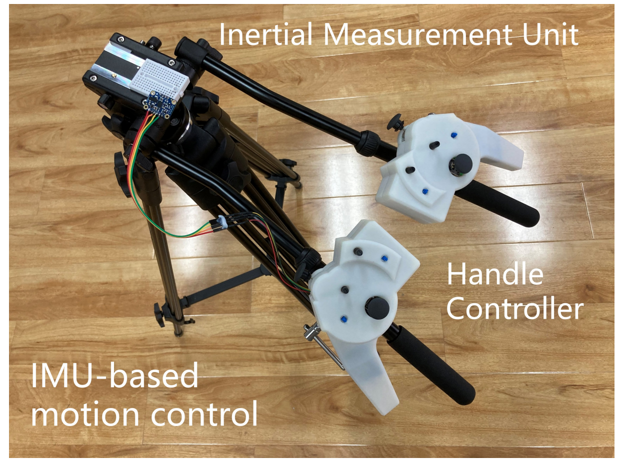
4.1 Hardware Design

Our VR training system, *VAction*, runs on a computer, allowing various project requirements and configurations to be adjusted before entering the VR scene. For specific training needs, users can wear a VR headset to enter the virtual scene within the training system.

In addition to the basic hardware components required to run the system and display content, such as a computer, a VR headset, and a video capture card, we have specifically designed a pair of handheld controllers for physical interaction. In our interview discussions, the simulation of operational feel was identified as a critical requirement. However, existing research has rarely focused on this aspect, especially within VR systems. We believe that traditional interaction modes are insufficient to replicate the tactile experience of operating a camera. Conventional mouse and keyboard inputs, as well as gesture interactions in VR, differ fundamentally from the operational modes of cameras. Commonly used game controllers and VR controllers typically feature two joysticks and multiple buttons, which not only fail to encompass all the functionalities required for camera operation but also exhibit a distinct operational mode that diverges from the controls of a real camera. This limitation restricts their effectiveness in training scenarios. Consequently, we have designed a specialized controller that closely simulates the



(a) Functional diagram of virtual camera



(b) Example of using the IMU function of the controller

Figure 2: Virtual camera controller.

operational feel of an actual camera within a VR system through integrated hardware and software interactions.

This controller must encompass all degrees of freedom and adjustment functions of a real camera. Specifically, it includes five spatial dimensions (free movement along the XYZ axes and pitch and roll rotations) and zoom (all controlled by joysticks), focus, aperture, and damping adjustments (all controlled by knobs), as well as mode switching and additional functions (controlled by switches). To simulate the operational mode of a real camera, we designed the controller as a pair of symmetric left and right handles. Physical operations are converted into electrical signals through potentiometers and switches.

The controller communicates with the computer via the UDP protocol, which is chosen based on the consideration of low-cost consumption and fast processing speed, thereby ensuring high-frequency data exchange for real-time control of the virtual camera.

To ensure effective practice with this controller, its design focuses on replicating the operational logic experienced when interacting with real cameras. The challenge in this process lies in the difference in the operational logic of the various components. For instance, both focus and zoom adjustments are controlled by potentiometers that convert angles into voltage values. However, in real cameras, the focus value corresponds to the absolute rotational angle, whereas the zoom value changes relative to the offset from the center position. In our controller, we adhered fully to the real operational logic, using a knob for focus adjustments and a joystick for zoom adjustments, as shown in Fig. 2a. The implementation of functions that correspond entirely to the operational logic is what current VR controllers are unable to achieve.

To further enhance realism, the controller integrates an Inertial Measurement Unit (IMU) for motion detection, allowing control of the pitch and roll of the virtual camera as shown in Fig. 2b. The controller can be mounted on a tripod, and the user can freely adjust the counterweight to simulate the heavy operation feel of real equipment, offering a portable and flexible setup due to its lightweight, detachable design.

Figure 3: The user interface within the VR training system

Overall, the hardware design prioritizes realistic camera operation, cost-effectiveness, and portability, providing practical, accessible training.

4.2 Software Design

VR technology offers a cost-effective and efficient solution for traditional film-shooting education. To maximize its benefits, we implemented a lightweight design that enables quick, convenient practice tailored to specific needs. Users are immersed in a VR scene with a system interface, as shown in Fig. 3, which includes functional modules such as virtual camera display, scene management, lens management, and camera movement control.

To ensure immersion and effective training, the VR environment and camera views must align with real-world counterparts, requiring high-fidelity scenes and precise camera parameter matching. For fictional scenes, high-resolution models from platforms like the Unreal Marketplace are used. For real-world environments, two methods are provided: accurate laser scanning and quicker computer vision method, detailed in Section 4.3. After 3D reconstruction, digital assets like lighting are integrated via the environmental management module to simulate settings and enhance information presentation.

In addition to accurate scene reconstruction, simulating lens parameters is essential for matching virtual and real shooting images. Using Zhang's calibration method [63], distortion parameters for specific focal lengths can be calculated in Unreal Engine for specific focus distance. This allows us to simulate common lens in under an hour, reducing the need for costly professional equipment. Reconstructed scenes and calibrated lenses can be quickly loaded and switched as needed.

To ensure completeness of the training process in the VR system, we adopted a lightweight, modular design. This includes modules for scene and lens selection, virtual camera control, and shot recording. For example, follow focus functionality is integrated into the controller, jib arm and static shots can be switched in camera mode, and Steadicam effects are achieved via Unreal Engine Blueprints. This allows students to focus on specific skills, lowering the initial learning barrier and enabling a more manageable, targeted learning experience.

For camera movement training, we offer two methods: automatic generation of virtual camera scripts using keyframes, where students follow predefined camera movements, and the option for students to record their own virtual footage. By comparing their recordings with system-generated examples, students can identify and correct mistakes, improving their camera operation skills through iterative practice.

4.3 Implementation

Hardware implementation

The entire controller prototype is constructed using 3D printing and electronic components including an ESP32 chip[1], with a total cost of approximately $90 and a weight of approximately 400g. The details of each component of our specially designed controller are shown in Table 3. If mass production is undertaken, the cost of the components including the casing, fixtures, and chip modules will be further reduced significantly. Compared to professional equipments that cost thousands of dollars, our controller offers substantial advantages in both price and portability.

The circuit design and PCB layout for the left controller are shown in Fig. 4. The program running on the ESP32 chip is written and uploaded using Arduino. Its primary function is to connect to the computer running the system via the UDP protocol, sending the voltage values of each component of the controller in JSON format. Testing shows a delay of about 30 ms, which is almost imperceptible. Our controller is able to provide near real-time feedback for user input.

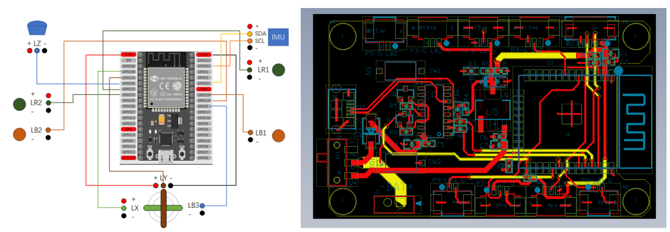
Software implementation

The VR program of the system runs on Unity and is streamed and displayed on the Oculus Quest 2 VR headset via Oculus Link. The main interface is implemented based on UE, with a plugin[2] in Unity that displays the UE system interface within the virtual scene. This approach allows the system to be used outside of VR, enabling teachers to quickly preview the teaching process.

[1]https://www.espressif.com.cn/products/socs/esp32
[2]https://github.com/hecomi/uWindowCapture

Table 3: Details of the components used in the controller

Component Name	Quantity	Price ($)	Weight (g)
3D Printed Controller Shell	1	30.00	50
Crab Claw Clamps	2	6.50	130
ESP32 Chips	2	5.00	10
Potentiometers	4	0.50	<5
Knobs	4	0.50	<5
Switches	4	0.50	<5
IMU Module	1	25.00	10
Lithium Battery	2	2.50	30



(a) The circuit design diagram for the left controller **(b) The PCB layout for the left controller**

Figure 4: The circuit design diagram and PCB layout for the pair of symmetrically designed controllers.

(a) Scene reconstructed through computer vision methods **(b) Refined digital scenes obtained through CAD modeling**

Figure 5: Example for Computer Vision Methods for Rapid Scene Modeling

The specific implementation of the two real-scene 3D reconstruction methods we provide is as follows:

Laser Scanning [56] for Accurate Scene Modeling: We acquire millimeter-level point cloud data of the scene from multiple angles using a laser scanner, and then complete the scene modeling using CAD (computer aided design) methods based on the spatial relationships found in the onsite photographs. Finally, we further refine the textures and materials on the model using these photographs.

Computer Vision Techniques for Rapid Scene Modeling: This lightweight method uses RGB-D data to quickly reconstruct a rough scene model [62]. A 2-minute video of the room shot with a smartphone is sufficient to create a basic model. As shown in Fig. 5, teachers can use this approach to preview the teaching environment, manually refine details through CAD modeling, and complete the process in under 2 hours.

The calibration and simulation of camera parameters are achieved through the plugin *Camera Calibration*[3] in UE, while the overall interface, module functionalities, and communication with the controller are all implemented using UE Blueprints and C++ code.

5 User Experiment

To verify the system's effectiveness in supporting participants' acquisition of camera skills, we conducted a user experiment including subjective scale ratings, behavioral experiment measurements, and theoretical knowledge evaluations. The experiment was designed based on a basic film-shooting lesson: operating the camera's Pan, Tilt, and Zoom functions.

The experiment was conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Research Ethics Committee of a local University. All participants provided written informed consent prior to participation.

5.1 Experimental Framework

The experiment comprised a 20-30 minute learning stage for skill acquisition and a 5-10 minute evaluation stage involving real camera operation and theoretical knowledge testing. Participants were randomly assigned to a control group and an experimental group, both of which are required to complete tasks in both the learning and evaluation stages. In the learning stage, the control group used traditional methods, that is, learning through videos and illustrations, while the experimental group trained with the VR film-shooting system. During the evaluation stage, both groups used a real professional camera to complete storyline-based filming tasks and filled out questionnaires to assess their theoretical knowledge.

If the virtual training system outperforms the traditional mode, it would demonstrate its superiority; if not, it would indicate a lower effectiveness.

5.2 Participants

This experiment recruited 24 beginners in film-shooting skills as participants (14 males and 10 females, all undergraduate or graduate students) from local universities with an average age of 23.52 ± 2.76 years. All participants have normal or corrected-to-normal vision, and did not experience 3D motion sickness during the experiment. Before the experiment, participants were briefed on the system's background, purpose, and process. They were taught basic operations and given time to familiarize themselves with the virtual or real system. After signing the informed consent form, the experiment would be conducted for approximately 40 minutes.

5.3 Experimental Design

During the learning stage, a 2 (learning mode: traditional, virtual) $\times$ 4 (level of difficulty-easy-to-hard: 1, 2, 3, 4) mixed design was conducted.

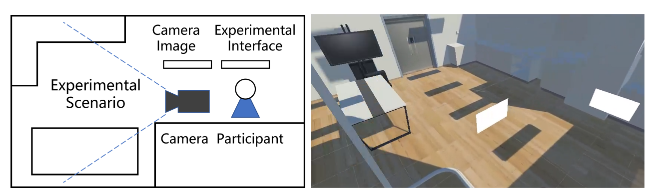
Learning mode is a between-subject factor with two levels: the traditional learning mode (control group) and the virtual learning mode using *VAction* (experimental group). Both groups received identical theoretical learning materials. The control group involved learning cinematography skills through videos, text, and images, while the experimental group used the VR system to replicate example videos.

The difficulty level is a within-subject factor with four levels (from easy to difficult): Level 1 (static shooting), Level 2 (dynamic shooting), Level 3 (dynamic combination), and Level 4 (basic film-shooting techniques). The learning materials were presented sequentially from easy to difficult, based on their level of difficulty. For each difficulty level, five segments were presented (randomly selected from the material library, with each segment appearing only once per participant's experiment, and not repeated). After completing the low-difficulty materials, the participant proceeded to the next difficulty level. The specific task flow is detailed in Section. 5.6.

In the evaluation stage, both groups complete identical tasks: using the emotional cues provided and the knowledge acquired in the learning stage, they film three short clips with appropriate techniques. For instance, when the emotional cue is "intense duel", participants should employ the whip pan technique. After shooting, all participants evaluate their theoretical knowledge through a questionnaire. The dependent variables were the participants' learning effectiveness and experience, including task completion similarity, teacher rating, attention and relaxation scores and subjective scale scores.

5.4 Experimental Environment and Apparatus

5.4.1 ***Experimental Environment.*** The top view of the laboratory layout is shown in Fig. 6a. In a laboratory room, the left half serves as the shooting range for the real shooting scene, with tables, chairs, and some equipment placed there. The right half is used to place experimental instruments for participants to learn and operate in the space.



(a) The top view of the real lab **(b) The virtual lab scene**

Figure 6: The left figure shows the physical lab setup: a camera in front of the participant, two monitors on a table to the right (left for the shooting scene, right for video stimuli), and the blue dashed lines indicating the camera's capture range. In the lower left of the diagram are tables and chairs, in the upper left is a low cabinet, and at the middle on the right is a door. The right figure shows the virtual lab, mirroring the physical layout, with virtual equipment in the corresponding positions of the real setup. The virtual shooting screen and experimental interface were placed in the positions originally occupied by the camera and the computer displaying the experimental interface, respectively.

[3] https://dev.epicgames.com/documentation/en-us/unreal-engine/camera-lens-calibration-overview

(a) The real laboratory scene

(b) The real lab and virtual scene blend at 0.5 transparency

(c) The virtual laboratory scene

Figure 7: Alignment effect between the real laboratory scenes and virtual scenes on spatial dimensions

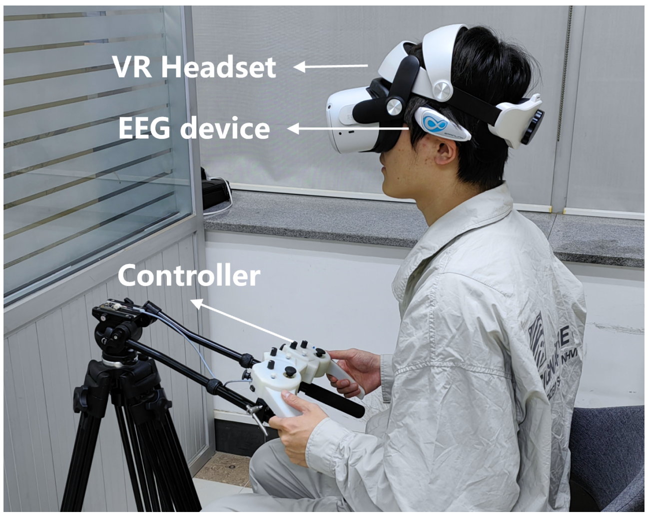


Figure 8: The participants in the experimental group wore Oculus Quest 2 VR headset and EEG device Brainlink Lite, and conducted the experiment using a self-developed controller.

To ensure consistency between virtual scenes and real shooting scenes, we used the laboratory as a prototype and constructed a virtual scene based on the 3D reconstruction method described in Section 4.2. This reconstruction accurately replicated the spatial dimensions of the laboratory as shown in Fig. 7. Efforts were made to match the furnishings in the laboratory as closely as possible. Virtual cameras and learning materials were placed at the same location as in the virtual laboratory, as shown in Fig. 6b. The dimensions and height from the ground of the hand controller device were kept consistent with those of a real camera. The goal is to facilitate the transfer of knowledge from the learning stage to the evaluation stage.

5.4.2 ***Experimental Apparatus.*** The real camera used was the Sony HDC1580 XDCAM camera, equipped with a Fujinon HA 18×7.6 BERD-S48 lens, connected to a CCU, and capable of transmitting captured footage to the host computer via a DeckLink Studio 4K video capture card. The VR film-shooting training system ran on a host computer with an Intel Core i9-12900K CPU and an ASUS ROG RTX 3090 Ti 24G Gaming graphics card. Virtual scenes and images are transmitted to the Oculus Quest 2 VR headset through Oculus Link, which has a monocular resolution of 1832 x 1920 pixels and a binocular resolution of 3664 x 1920 pixels, with a refresh rate of 90Hz. The hand controller, mounted on a tripod, communicated with the host computer via UDP protocol. A pedal was placed at the foot of the participants to interact with the experimental program that records data.

We used Brainlink's portable EEG device Brainlink-Lite[32] to dynamically record participants' EEG signal. The headband device includes a recording dry electrode, which is located at the Fp1 position on the participant's forehead according to the international 10-20 system. The EEG signal has a sampling rate of 512 Hz and is wirelessly transmitted through Bluetooth connection. The entire headband weighs no more than 50g, having virtually no impact on the participants' operations. The participant wearing a VR headset and the EEG device Brainlink-Lite during the experiment is shown in Fig. 8.

5.5 Experimental Stimuli

For the tasks in the learning stage, we developed a library of learning materials for shooting based on expert guidance, which includes both video and text resources. The video materials are categorized into theoretical knowledge introduction videos and example videos in both real and virtual scenes. The theoretical knowledge introduction video is approximately 2 minutes long. For each type of scene, there are four levels of difficulty, each level containing 10 videos, totaling 40 learning materials. The duration of each example video ranges from 7 to 15 seconds, with an average length of 10 seconds. Each difficulty level is accompanied by an explanatory text on conceptual knowledge (concepts and theories), ranging from 50 to 200 words. Each video within a difficulty level demonstrates an operational example, paired with a description of procedural knowledge (operational skills), consisting of 30 to 50 words. All emotional cues given in the evaluation stage originate from the conceptual knowledge learned in the learning stage.

5.6 Experimental Procedure

Upon arrival at the lab, participants were briefed on the tasks and signed an informed consent form. They then sat in front of the equipment and wore the Brainlink-Lite EEG device and Quest 2 VR

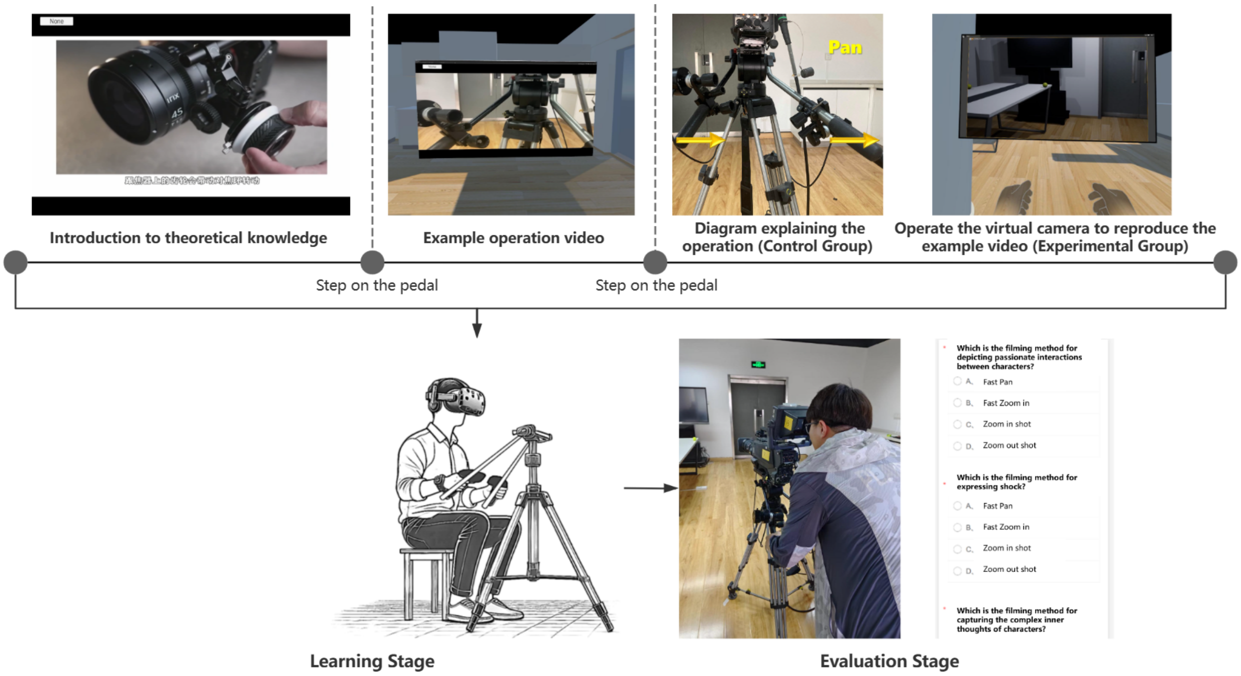


Figure 9: The experiment was divided into two stages. In the learning stage, participants learned theoretical knowledge and operational skills separately. In the evaluation stage, participants' learning outcomes were assessed through the results of shooting with real cameras and an online questionnaire.

headset before being randomly assigned to either the traditional or virtual learning group.

The experiment was divided into two stages: learning stage and evaluation stage. The experimental flowchart experienced by each participant is shown in Fig. 9.

5.6.1 ***Learning Stage***. In the subsequent learning stage, participants put on the Oculus Quest 2 VR headset and entered the virtual shooting training system, first watching a theoretical knowledge introduction video. They then engaged in free practice to familiarize themselves with the learning task process before beginning the formal learning stage. Learning materials were presented in order of increasing difficulty. For each difficulty level, participants first viewed the corresponding text and images of conceptual knowledge. They then randomly selected five video segments from the material library to practice the relevant skills. After completing the five segments, they progressed to the next difficulty level.

The experimental interface is shown in Fig. 10a. The virtual screen displayed the learning materials that both the control group and the experimental group view consistently, including illustrated explanations of theoretical knowledge and operational demonstration videos. After watching each operational video, the control group would continue to view the corresponding illustrated explanation of the operation, while the experimental group would replicate the actions from the examples by controlling the virtual camera view in *VAction* using the controller. The virtual camera view is presented on the left screen as shown in Fig. 10b. They used a foot pedal to start and stop recording after each task. A 5-second blank screen marked the transition to the next task.

(a) Experimental interface in VR **(b) Virtual camera images in VR**

Figure 10: During the learning stage, participants were presented with an experimental interface as shown in (a), displaying theoretical knowledge and operational demonstration videos. On the left side of the experimental group participants, there was an additional screen as shown in (b), showing the virtual camera view from *VAction*. Participants in the experimental group used a controller to manipulate this view and replicate the demonstrated operations.

5.6.2 ***Evaluation Stage***. After completing the learning tasks, participants removed the VR headset and EEG device, took a short break, and moved to the evaluation stage. Here, they received randomly assigned emotion cues (such as "intense duel") and used real professional cameras to film 20-second clips. The subjects of these clips were consistent with those used in the learning stage. Each participant received three emotional cues, which they used to create three corresponding short films. The filming of these short films required them to master the correct theoretical filming techniques while also performing smooth operations. Subsequently, two film production teachers with over eight years of teaching experience evaluated the accuracy of expression (whether the correct filming techniques were employed) and smoothness of operation in these short films.

Finally, participants completed a questionnaire consisting of eight multiple-choice questions assessing their retention of theoretical knowledge from the learning stage. These scores reflected their understanding of the material.

5.7 Data Collection

In the experiment, data collection included both learning performance and system experience. Learning performance was measured by attention scores, relaxation scores, replication similarity for the experimental group, and teacher ratings of the short films. System experience was assessed through usability questionnaires during the learning process.

Performance during the learning stage: Attention and relaxation scores during the learning stage were measured using the Brainlink-Lite EEG device, which provided reference values every second via Bluetooth at a frequency of 1 Hz. Participants wore the device throughout the learning stage to collect EEG data for attention and relaxation. We chose the device because it was widely used in practical human-machine interactive systems [7, 31, 43, 57].

The operational results of the learning stage: The replication similarity score for the experimental group was derived by weighting the similarity of the first and last frames of the operation and the overall video similarity. The similarity of the first and last frames was calculated using the Structural Similarity Index (SSIM), while the overall video similarity was assessed by calculating the Hamming distance between proportionally sampled frames.

The operational results of the evaluation stage: The short films produced during the evaluation stage were assessed by two film production teachers. They scored the films based on the accuracy of expression and the smoothness of operation.

Subjective experience during the learning stage: Participants completed the third version of Post-study System Usability Questionnaire (PSSUQ) [54] to assess the overall user experience with the system's usability.

Mastery of knowledge: After completing the storyboard based filming task, each participant received an online questionnaire with 8 multiple-choice questions to assess their mastery of theoretical knowledge.

5.8 Results

Attention and relaxation scores: The learning stage was divided into theoretical learning phase and skill practice phase, with the EEG data separated by participants pressing a foot pedal to mark the transition between these two phases. We conducted 2×2 ANOVA analysis on attention and relaxation scores respectively as shown in Fig. 11. The attention scores during the skill practice phase (46.8) were significantly higher than those during the theoretical learning phase (43.2), $F_{(1,176)} = 5.281$, $p < 0.05$, ${\eta_p}^2 = 0.029$. The relaxation scores in the experimental group using *VAction* (58.7) were significantly higher than those in the control group (54.5) during both phases, $F_{(1,176)} = 13.756$, $p < 0.001$, ${\eta_p}^2 = 0.072$. In addition, it showed a trend that the control group showed higher attention scores during the theoretical learning phase, whereas the experimental group showed higher attention scores during the skill practice phase.

The replication similarity score: In the experimental group, participants replicated example videos during skill practice. We calculated the average similarity of the replication videos across different difficulty levels for each participant. After removing missing data (three participants whose operation videos failed to save, making similarity calculations impossible), the variation curve (as shown in Fig. 12) shows that as difficulty increases, the similarity of replications decreases. However, after practice, participants achieved basic mastery, with an average similarity close to 75%, and performance at the highest difficulty level similar to the third level.

Score for short films based on storylines: Independent samples t-tests were conducted on the standardized scores across the groups, as shown in Fig. 13. The experimental group using *VAction* had a significantly higher score for the expression accuracy (62.9/100) than the control group utilizing traditional learning methods (49.4/100), $t(70) = 2.367$, $p < 0.05$, ${\eta_p}^2 = 0.074$. Additionally, the experimental group also had a significantly higher average score for operational smoothness (54.9/100) compared to the control group (41.0/100), $t(70) = 2.434$, $p < 0.05$, ${\eta_p}^2 = 0.078$.

Score of knowledge mastery questionnaire: The experimental group had an average score (49.2/80) on the theoretical knowledge questionnaire that was slightly lower than that of the control group (52.5/80). A simple effects t-test showed that $p = 0.50 > 0.05$, indicating that the difference in the level of mastery of theoretical knowledge between the control group and the experimental group was not significant.

The subjective questionnaires: The descriptive statistical results of participants' scores on various indicators such as system usefulness, information quality, interface quality, and overall usability are shown in Table 4. Comparing the scores obtained by the system with the norm reference scores for each indicator, it can be seen that our training system *VAction* has scores far higher than the reference scores, indicating that the system has good usability.

Table 4: The score of PSSUQ

Observation Indicators	Reference	System Score	Result
PSSUQ Overall usability	3.02	5.59	high
PSSUQ System usefulness	3.02	5.79	high
PSSUQ Infomation quality	3.24	5.35	high
PSSUQ Interface quality	2.71	5.63	high

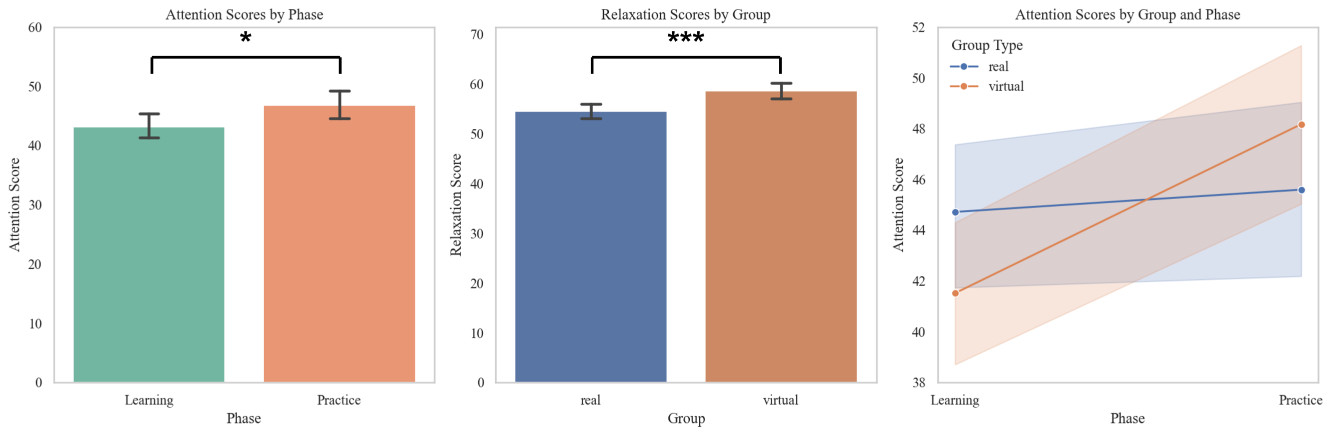


Figure 11: EEG data analysis results: participants' attention was significantly higher during the skill practice phase, and the experimental group showed better relaxation scores compared to the control group. Additionally, there was a trend where the experimental group focused more on the skill practice phase than the control group.

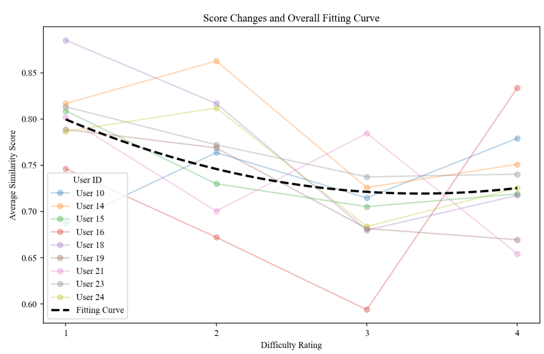


Figure 12: The curve of the average similarity between the operations reproduced by the experimental group and the examples

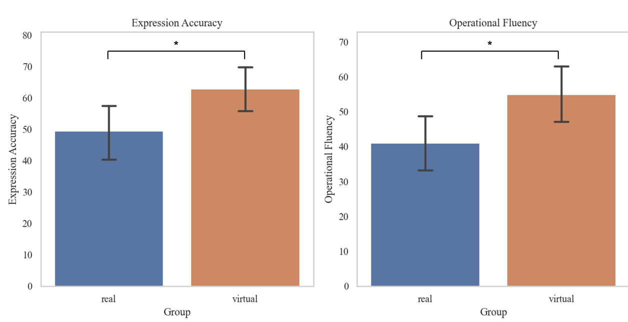


Figure 13: Analysis of average score results for short films

5.9 Post Experiment Interviews

After the experiment, we randomly invited three participants from the experimental group for interviews to discuss their experiences with *VAction*. They all agreed that *VAction* was more flexible and user-friendly than real cameras, with real cameras being notably heavier. One liked the realism of the weight, while two preferred the light virtual controllers, which encouraged them to engage more freely, stating that "It's like playing a game, without the worry of damaging expensive equipment." They also mentioned that the interface of the virtual system is clearer and free from irrelevant distractions, helping them concentrate on system prompts and practice results. On effectiveness, one noted that while VR helped understand basic operations, precise control was still challenging, and a half-hour session was too short to assess long-term training effects. Another suggested that VR's advancement could revolutionize film production and education, envisioning a fully VR-based production pathway and recommending exploration of machine-less film processes alongside simulating teaching workflows.

6 Discussion

6.1 Analysis and Discussion of Experimental Results

First, during the learning stage, we found that the relaxation scores of participants read by the EEG headband indicate that our virtual training system can effectively enhance the relaxation of beginners. VAction reduces the tension that novice users experience when operating complex equipment, making it easier for them to master skills efficiently. In terms of attention, experimental results show that beginners generally choose to focus more attention on the skill practice phase, which aligns with our interview findings.

Additionally, we observed a trend that beginners using *VAction* tend to concentrate more on the skill practice phase than traditional beginners. In traditional learning methods, students primarily acquire theoretical and operational knowledge through the teacher's explanations and demonstrations. This nonparticipatory learning approach results in no significant differences in the attention allocated during the theoretical learning phase and the skills practice phase. In contrast, when learning with *VAction*, students tend to shift their attention more from the relatively simple theoretical learning to the skill practice phase, mastering both theory and operation more effectively through a "learning-by-doing" approach.

The results have shown that the control group demonstrated a higher level of attention during the theoretical learning phase. This difference may stem from the fact that the experimental group, anticipating the subsequent practical training, allocated some cognitive resources to skill-practicing planning, while participants in the control group were able to fully concentrate on the video-based learning. However, questionnaire assessments in the evaluation stage revealed no significant difference between the two groups in terms of theoretical knowledge mastery. More importantly, the experimental group devoted more attention to the skill practice phase and ultimately outperformed the control group in both expression accuracy and operational smoothness during practical tasks, which validates the effectiveness of our training system in enhancing practical skills.

In the experiment, to eliminate the impact of VR on learning outcomes and to focus on exploring the learning effects of the *VAction* learning model, which combines software and hardware for a hands-on learning experience, we transplanted the traditional learning content of the control group into VR. However, the virtual scenes provided by the experiment, which are consistent with real shooting scenarios, inherently offer consistency in the learning context. This makes the learning transfer effect in VR superior to that of traditional learning methods, which struggle to provide professional shooting environments. In the process of film production teaching, the knowledge imparted includes conceptual knowledge (theory of cameras and shooting techniques, etc.) and procedural knowledge (how to operate a camera) [5]. As highlighted by the teacher, procedural knowledge is key in this practical discipline. In traditional learning mode, students need to transfer from conceptual knowledge to procedural knowledge, whereas the hardware of *VAction*, specifically the controller designed to replicate the operational feel of a real camera, provides an embodied operating experience, which is not available in previous VR film-shooting training systems. This enables students to easily acquire procedural knowledge and enhances the transfer of learning to real operations [33].

6.2 Extensible Applications

We have explored other applications for the system. *VAction* has been currently used at a local television station for pre-visualization prior to official shoots. Traditional pre-visualization requires complex set-ups and the presence of various department personnel for coordination. *VAction* allows for rapid scene switching, screen transitions, and lens adjustments, enabling staff to familiarize themselves with the process via computer without needing to be physically present on set, significantly increasing the efficiency of pre-visualization. Using the Faro Focus 3D X330 laser scanner, we captured precise point cloud data of a 400-square-meter broadcasting studio with 22 scanning stations over 5 hours. The scene was modeled with an error margin of less than 5 cm. The feature to generate exemplar camera movements using keyframes also provides professional camera operators with a more convenient way to preview shot effects. However, feedback from TV station professionals indicates that *VAction* is mainly used for pre-visualization visual effects visualization due to the need for real-time accuracy, with actual filming still done on real sets. In the actual industrial processes of TV broadcasting or film production, the focus is on the efficiency, yet most current VR systems do not possess the accuracy required for real-time broadcasting, thus they are primarily used for previews. In contrast, in film shooting education, we place a high value on the training feel and real-time feedback provided to students by VR systems. Therefore, we have designed this integrated software-hardware system specifically to address these aspects. Though not used in live shooting, this feedback has been encouraging and enlightening: *VAction*, aimed at beginners, offers a consistent operational practice experience, while in professional settings, its quick preview feature is valued for efficiency.

Additionally, the hardware controller has been successfully tested in other applications, such as racing games, showcasing its versatility and extensibility.

6.3 Limitations and Future Work

During the user experiment, we found that the training method of following example videos in *VAction* did not fully leverage the advantages of the VR system. When faced with complex operations, users may feel overwhelmed and focus more on replicating the actions themselves rather than understanding the application scenarios and corresponding theories. To address this limitation, we believe that existing teaching examples can be presented in a richer three-dimensional format within the VR environment, using agents for face-to-face instruction instead of merely watching videos. Additionally, users should have the option to pause at any time to view the details of the example operations from all angles.

To ensure portability, we adopted a lightweight design for the controller, which results in a feel difference when used alone compared to a real camera. We recommend mounting it on a tripod and adjusting the operational feel by altering the counterweights on the tripod's handle.

Our current VR teaching scenarios primarily replicate existing studios and classrooms. In reality, we could incorporate more classic scenes from films or recreate specific iconic moments related to the content being taught [39], thus integrating the learning of conceptual knowledge and procedural knowledge. These efforts will require detailed collaboration and guidance from more teachers with professional teaching experience.

7 Conclusion

In this paper, we presented *VAction*, a VR film-shooting training system. Through research and interviews with students and teachers in the field, we identified the challenges faced by traditional film production education and established the core principles for designing a VR system to address these challenges. Our system adheres to the principle of lightweight design, integrating high-fidelity virtual scenes and shots with a self-developed controller, accurately replicating the operational processes and tactile feel of a real camera. In a user experiment involving 24 beginners in film-shooting, we found that *VAction* effectively engages users' attention by providing immersive, high-fidelity scenes and embodied experiences that replicate real operations. This allows users to focus more on the crucial phase of acquiring procedural knowledge and significantly improves their learning outcomes. These findings offer greater possibilities for film production education and support the development of similar VR skill training systems.

Acknowledgments

We would like to thank our participants for volunteering their time and insights and our anonymous reviewers for their thoughtful feedback and suggestions on improving this work. We thank the support from the MOE-Key Laboratory of Pervasive Computing. The work was partially supported by the Natural Science Foundation of China (62332019, U2336214, 62461160309) and the Research Project of the State Key Laboratory of Ultra HD Video and Audio Production and Broadcasting Presentation of China Media Group (CMGSKL2021KF015).

References

[1] Gokce Akcayir and Carrie Demmans Epp (Eds.). 2021. *Designing, Deploying, and Evaluating Virtual and Augmented Reality in Education.* IGI Global. https://doi.org/10.4018/978-1-7998-5043-4

[2] Emrah Akman and Recep Çakır. 2019. Pupils' Opinions on an Educational Virtual Reality Game in Terms of Flow Experience. *International Journal of Emerging Technologies in Learning* 14, 15 (2019).

[3] Víctor Alejandro Huerta-Torruco, Óscar Hernández-Uribe, Leonor Adriana Cárdenas-Robledo, and Noé Amir Rodríguez-Olivares. 2022. Effectiveness of virtual reality in discrete event simulation models for manufacturing systems. *Computers & Industrial Engineering* 168 (2022), 108079. https://doi.org/10.1016/j.cie.2022.108079

[4] John R Anderson. 1996. ACT: A simple theory of complex cognition. *American psychologist* 51, 4 (1996), 355.

[5] Lorin W Anderson and David R Krathwohl. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition.* Addison Wesley Longman, Inc.

[6] Steven Ascher and Edward Pincus. 2007. *The filmmaker's handbook: A comprehensive guide for the digital age.* Penguin.

[7] Yulong Bian, Chao Zhou, Yang Zhang, Juan Liu, Jenny Sheng, and Yong-Jin Liu. 2023. Focus on Cooperation: A Face-to-Face VR Serious Game for Relationship Enhancement. *IEEE Transactions on Affective Computing* (2023), 1–16. https://doi.org/10.1109/TAFFC.2023.3306198

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa

[9] Mihaly Csikszentmihalyi. 1997. Flow and the psychology of discovery and invention. *HarperPerennial, New York* 39 (1997), 1–16.

[10] John Dewey. 1986. Experience and education. In *The educational forum*, Vol. 50. Taylor & Francis, 241–252.

[11] David Dickson and AVID BAMFORD. 1995. Improving the interpersonal skills of social work students: The problem of transfer of training and what to do about it. *The British Journal of Social Work* 25, 1 (1995), 85–105.

[12] Jane L. E, Ohad Fried, Jingwan Lu, Jianming Zhang, Radomír Mech, Jose Echevarria, Pat Hanrahan, and James A. Landay. 2020. Adaptive Photographic Composition Guidance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376635

[13] Andres Forero-Serna. 2024. Toward the Consolidation of Filming Simulators: A Systematic Literature Review of Virtual Environments for Audiovisual Learning and Practice. *International Journal of Human–Computer Interaction* 0, 0 (2024), 1–14. https://doi.org/10.1080/10447318.2024.2383035 arXiv:https://doi.org/10.1080/10447318.2024.2383035

[14] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 662, 14 pages. https://doi.org/10.1145/3411764.3445576

[15] Jens Grubert, Eyal Ofek, Michel Pahud, and Per Ola Kristensson. 2018. The Office of the Future: Virtual, Portable, and Global. *IEEE Computer Graphics and Applications* 38, 6 (2018), 125–133. https://doi.org/10.1109/MCG.2018.2875609

[16] David Hamilton, Jim McKechnie, Edward Edgerton, and Claire Wilson. 2021. Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education* 8, 1 (2021), 1–32.

[17] Gert Jan Harkema and André Rosendaal. 2020. From cinematograph to 3D model: how can virtual reality support film education hands-on? *Early Popular Visual Culture* 18, 1 (2020), 70–81. https://doi.org/10.1080/17460654.2020.1761598 arXiv:https://doi.org/10.1080/17460654.2020.1761598

[18] George E Hein. 1991. Constructivist learning theory. *Institute for Inquiry. Available at:/http://www. exploratorium. edu/ifi/resources/constructivistlearning. htmlS* (1991).

[19] Dong Huang and Jong Hoon-Yang. 2023. Artificial Intelligence Combined with Deep Learning in Film and Television Quality Education for the Youth. *International Journal of Humanoid Robotics* 20, 06 (2023), 2250019. https://doi.org/10.1142/S0219843622500190 arXiv:https://doi.org/10.1142/S0219843622500190

[20] Hsiu-Mei Huang and Shu-Sheng Liaw. 2018. An analysis of learners' intentions toward virtual reality learning based on constructivist and technology acceptance approaches. *International review of research in open and distributed learning* 19, 1 (2018).

[21] Susan Jang, Jonathan M Vitale, Robert W Jyung, and John B Black. 2017. Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education* 106 (2017), 150–165.

[22] Hongda Jiang, Marc Christie, Xi Wang, Libin Liu, Bin Wang, and Baoquan Chen. 2021. Camera keyframing with style and control. *ACM Trans. Graph.* 40, 6, Article 209 (Dec. 2021), 13 pages. https://doi.org/10.1145/3478513.3480533

[23] Hongda Jiang, Bin Wang, Xi Wang, Marc Christie, and Baoquan Chen. 2020. Example-driven virtual cinematography by learning camera behaviors. *ACM Trans. Graph.* 39, 4, Article 45 (Aug. 2020), 14 pages. https://doi.org/10.1145/3386569.3392427

[24] Qiao Jin, Yu Liu, Svetlana Yarosh, Bo Han, and Feng Qian. 2022. How Will VR Enter University Classrooms? Multi-stakeholders Investigation of VR in Higher Education. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 563, 17 pages. https://doi.org/10.1145/3491102.3517542

[25] Qiao Jin, Yu Liu, Ye Yuan, Bo Han, Feng Qian, and Svetlana Yarosh. 2024. Virtual Reality, Real Pedagogy: A Contextual Inquiry of Instructor Practices with VR Video. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 665, 21 pages. https://doi.org/10.1145/3613904.3642510

[26] Mina C Johnson-Glenberg. 2018. Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI* 5 (2018), 375272.

[27] Janine Jongbloed, Rawad Chaker, and Elise Lavoué. 2024. Immersive procedural training in virtual reality: A systematic literature review. *Computers & Education* (2024), 105124.

[28] David A Kolb. 2014. *Experiential learning: Experience as the source of learning and development.* FT press.

[29] Yanni Liao. 2024. Education in cinematography and VR- technologies: the impact of animation on the film perception. *Interactive Learning Environments* 0, 0 (2024), 1–15. https://doi.org/10.1080/10494820.2024.2372831 arXiv:https://doi.org/10.1080/10494820.2024.2372831

[30] Robb Lindgren, Michael Tscholl, Shuai Wang, and Emily Johnson. 2016. Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & education* 95 (2016), 174–187.

[31] Tianren Luo, Gaozhang Chen, Yijian Wen, Pengxiang Wang, yachun fan, Teng Han, and Feng Tian. 2024. Exploring the Effects of Sensory Conflicts on Cognitive Fatigue in VR Remappings. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 53, 16 pages. https://doi.org/10.1145/3654777.3676439

[32] Macrotellect. 2023. Brainlink. https://o.macrotellect.com/default.html.

[33] Guido Makransky, Stefan Borre-Gude, and Richard E Mayer. 2019. Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning* 35, 6 (2019), 691–707.

[34] Guido Makransky and Gustav B Petersen. 2021. The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review* 33, 3 (2021), 937–958.

[35] Information Resources Management Association. 2019. *Virtual Reality in Education: Breakthroughs in Research and Practice.* IGI Global. https://doi.org/10.4018/978-1-5225-8179-6

[36] Anthony Marini and Randy Genereux. 2013. The challenge of teaching for transfer. In *Teaching for transfer*. Routledge, 1–19.

[37] Andreas Marougkas, Christos Troussas, Akrivi Krouska, and Cleo Sgouropoulou. 2023. Virtual Reality in Education: A Review of Learning Theories, Approaches and Methodologies for the Last Decade. *Electronics* 12, 13 (2023). https://www.mdpi.com/2079-9292/12/13/2832

[38] Alan Medlar, Mari Tatsuko Lehtikari, and Dorota Glowacka. 2024. Behind the Scenes: Adapting Cinematography and Editing Concepts to Navigation in Virtual Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 545, 12 pages. https://doi.org/10.1145/3613904.3642412

[39] Brittany Morago and David Monahan. 2021. Enhancing Film Education Learning Outcomes With Virtual Experiences. *IEEE Computer Graphics and Applications* 41, 2 (2021), 99–105. https://doi.org/10.1109/MCG.2021.3050949

[40] Thomas Muender, Thomas Fröhlich, and Rainer Malaka. 2018. Empowering Creative People: Virtual Reality for Previsualization. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188612

[41] Thomas Muender, Anke V. Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. 2019. Does It Feel Real? Using Tangibles with Different Fidelities to Build and Explore Scenes in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300903

[42] Thomas Muender, Georg Volkmar, Dirk Wenig, and Rainer Malaka. 2019. Analysis of Previsualization Tasks for Animation, Film and Theater. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312953

[43] Geeta U Navalyal and Rahul D Gavas. 2014. A dynamic attention assessment and enhancement tool using computer graphics. *Human-centric Computing and Information Sciences* 4 (2014), 1–7.

[44] Matthias Oberhauser and Daniel Dreyer. 2017. A virtual reality flight simulator for human factors engineering. *Cognition, Technology & Work* 19 (2017), 263–277.

[45] Prajakt Pande, Amalie Thit, Anja Elaine Sørensen, Biljana Mojsoska, Morten E Moeller, and Per Meyer Jepsen. 2021. Long-Term Effectiveness of Immersive VR Simulations in Undergraduate Science Learning: Lessons from a Media-Comparison Study. *Research in Learning Technology* 29 (2021).

[46] Mitch R. Paro, David S. Hersh, and Ketan R. Bulsara. 2022. History of Virtual Reality and Augmented Reality in Neurosurgical Training. *World Neurosurgery* 167 (2022), 37–43. https://doi.org/10.1016/j.wneu.2022.08.042

[47] Gustav Bøg Petersen, Giorgos Petkakis, and Guido Makransky. 2022. A study of how immersion and interactivity drive VR learning. *Computers & Education* 179 (2022), 104429. https://doi.org/10.1016/j.compedu.2021.104429

[48] Stéphanie Philippe, Alexis D. Souchet, Petros Lameras, Panagiotis Petridis, Julien Caporal, Gildas Coldeboeuf, and Hadrien Duzan. 2020. Multimodal teaching, learning and training in virtual reality: a review and case study. *Virtual Reality & Intelligent Hardware* 2, 5 (2020), 421–442. https://doi.org/10.1016/j.vrih.2020.07.008 VR/AR research and commercial applications in Singapore.

[49] Johanna Pirker, Isabel Lesjak, Mathias Parger, and Christian Gütl. 2018. An educational physics laboratory in mobile versus room scale virtual reality-a comparative study. In *Online Engineering & Internet of Things: Proceedings of the 14th International Conference on Remote Engineering and Virtual Instrumentation REV 2017, held 15-17 March 2017, Columbia University, New York, USA*. Springer, 1029–1043.

[50] Giuseppe Riva, Fabrizio Davide, and Wijnand A IJsselsteijn. 2003. Being there: The experience of presence in mediated environments. *Being there: Concepts, effects and measurement of user presence in synthetic environments* 5 (2003), 2003.

[51] Gavriel Salomon and David N Perkins. 1992. Transfer of learning. *Computer Science* (1992).

[52] Ganesh Sankaranarayanan, Lizzy Wooley, Deborah Hogg, Denis Dorozhkin, Jaisa Olasky, Sanket Chauhan, James W Fleshman, Suvranu De, Daniel Scott, and Daniel B Jones. 2018. Immersive virtual reality-based training improves response in a simulated operating room fire scenario. *Surgical endoscopy* 32 (2018), 3439–3449.

[53] Mike Sinclair, Eyal Ofek, Christian Holz, Inrak Choi, Eric Whitmire, Evan Strasnick, and Hrvoje Benko. 2018. Three Haptic Shape-Feedback Controllers for Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 777–778. https://doi.org/10.1109/VR.2018.8446399

[54] Tom Tullis and Bill Albert. 2013. Chapter 6 - Self-Reported Metrics. In *Measuring the User Experience (Second Edition)* (second edition ed.), Tom Tullis and Bill Albert (Eds.). Morgan Kaufmann, Boston, 121–161. https://doi.org/10.1016/B978-0-12-415781-1.00006-6

[55] Georg Volkmar, Thomas Muender, Dirk Wenig, and Rainer Malaka. 2020. Evaluation of Natural User Interfaces in the Creative Industries. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3375201

[56] Jing Wang, Juan Zhang, and Qingtong Xu. 2014. Research on 3D laser scanning technology based on point cloud data acquisition. In *2014 International Conference on Audio, Language and Image Processing*. 631–634. https://doi.org/10.1109/ICALIP.2014.7009871

[57] Jiahui Xu and Baichang Zhong. 2018. Review on portable EEG technology in educational research. *Computers in Human Behavior* 81 (2018), 340–349.

[58] Ruyu Yan, Jiatian Sun, Longxiulin Deng, and Abe Davis. 2022. ReCapture: AR-Guided Time-lapse Photography. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 36, 14 pages. https://doi.org/10.1145/3526113.3545641

[59] Li Ying, Zhang Jiong, Sun Wei, Wang Jingchun, and Gao Xiaopeng. 2017. VREX: Virtual reality education expansion could help to improve the class experience (VREX platform and community for VR based education). In *2017 IEEE frontiers in education conference (FIE)*. Ieee, 1–5.

[60] Zhiyuan Yu, Cheng-Hung Lo, Mutian Niu, and Hai-Ning Liang. 2023. Comparing Cinematic Conventions through Emotional Responses in Cinematic VR and Traditional Mediums. In *SIGGRAPH Asia 2023 Technical Communications* (Sydney, NSW, Australia) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 10, 4 pages. https://doi.org/10.1145/3610543.3626175

[61] Mark M. Zaki, Rushikesh S. Joshi, Jacob R. Joseph, Yamaan S. Saadeh, Osama N. Kashlan, Jakub Godzik, Juan S. Uribe, and Paul Park. 2024. Virtual Reality–Enabled Resident Education of Lateral-Access Spine Surgery. *World Neurosurgery* 183 (2024), e401–e407. https://doi.org/10.1016/j.wneu.2023.12.108

[62] Jiazhao Zhang, Chenyang Zhu, Lintao Zheng, and Kai Xu. 2021. ROSEFusion: Random Optimization for Online Dense Reconstruction under Fast Camera Motion. *CoRR* abs/2105.05600 (2021). arXiv:2105.05600 https://arxiv.org/abs/2105.05600

[63] Z. Zhang. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 11 (2000), 1330–1334. https://doi.org/10.1109/34.888718

[64] Yuhang Zhao, Cynthia L. Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 2018. Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173690

[65] Minhua Ma Zhejun Liu, Yunshui Jin and Jiachen Li. 2023. A Comparison of Immersive and Non-Immersive VR for the Education of Filmmaking. *International Journal of Human–Computer Interaction* 39, 12 (2023), 2478–2491. https://doi.org/10.1080/10447318.2022.2078462 arXiv:https://doi.org/10.1080/10447318.2022.2078462