Proceedings of ICLLCD 2026 Symposium: Intelligent Media for Cultural Bridge: Forum on Global-Local Communication
DOI: 10.54254/2753-7064/2026.BJ33608

One Shot into the Scene: Gamified Interaction in VR Filmmaking Process

Siqi Tong[1], Lin Zhu[2*]

[1]*Department of Imaging Design and Digital Art, Hong Kong Metropolitan University, Hong Kong SAR, China*
[2]*Institute of Art and Design, Beijing City University, Beijing, China*
**corresponding Author. Email: 202405006@bucm.edu.cn*

Abstract. The digital age is witnessing the deep integration of artificial intelligence and virtual reality, and generative AI, VR agents, as well as embodied interaction technologies have brought forth new film formats—this comes as a response to the high thresholds associated with traditional film production. Integrating interactive narrative theory with the gamified design philosophy of VR, this research establishes a two-stage full-process interactive system, which is composed of image construction and image deconstruction. This study realizes interactive companionship within the game by converting the key procedures of film production into a game and inserting intelligent agents to take the place of conventional fixed NPCs. There is a specific requirement for the first level: players must finish filming the scene, whereas the second level uncovers the inherent logic of image production via embodied interaction. Not only does this research align with the developmental direction of digital content creation in the AI era, but it also offers a novel approach and practical guidance for improving the media literacy of the general public and innovating the interactive modes of VR film and television.

Keywords: VR film, Gamification interaction, VR Agent, Immersive creation, Serious Game Design

1. Introduction

Now comes the Digital Age, people take a look at where A I hits its peak and meets VR, here is what people call a big mixing process - coming along are generational AI's the VR agents and embodied interactions creating all new form movie for movies that can respond to those high production barriers and by adding narrative theory with the game – like ideas within the own VR technology. Putting everything back together again and producing a completely interactive procedure kind of thing with two simple steps - building an image, breaking down an image

This research has turned the key processes in making films into games and added intelligent agents to play as NPCs that are usually set in stone so people can have interactive companions in the game. For the first one, they have to finish recording the scene, and then on the second level it's about figuring out the internal logic of picture making through embodiment. Not only is it in line

DOI: 10.54254/2753-7064/2026.BJ33608

with the development direction of digital content production in the AI era, but also it gives a new way and concrete help to enhance the media literacy of ordinary people and innovate the interaction methods of VR film and television. And generative AI and VRagents along with Embodied Interaction tech has broken down all professional barriers and people had a lot of different kinds of experiences which i didn't get from making a ordinary film. It's also made it possible for people to do smart guidance as well as give personalized talkback in an artificial environment. The era of AI&VR has brought great strength, people do not just want to watch movies, they want to participate in all aspects of creation, and understand how a movie is made from the start to finish [1].

And ordinary people can't reach the main stuff like writing scripts, arranging shots and using audio-visual languages. Most of the current VR films and TV works, as well as creation oriented interactive content are still in a one way or segmented experience, not fully replicating all aspects of making a film. There aren't many serious games right now in the vr space that were made for film. More so, most of the interactive character on VR games are just a NPC and they can't be used as personal companion [2] That's why I'm going to make it a serious interactive game where the VR intelligent agent is focused on making films.

In this game there's a film agent who works as a game guide, so it can break away from the limitation of old mechanical interaction and becomes an important interactive medium during all stages of image building and deconstruction [3, 4]. This is an agent that acts like a guide who can talk. Pre-image construction, the film agent gives out personalized advice right in real time according to what the player picks from story pieces and making shots, it changes with the player's ideas about the plot line and shooting thoughts. Post image disassembly phase - Film Agent becomes Media Reflection Guide: follow the players roaming paths and disassemblies and explain each steps of building up the pictures and give simple shot secrets [5]. So as to give them real time and relevant mental coaching during the dismantling phase. With the help of generative intelligence, the film agent has gone from being just a passive responder to an active companion.

In addition it can make the professional guidance for film production more humanized and adaptable, and also bring some flexibility and immersive to the whole process interaction, which strengthens the core design idea of "creation is a kind of experience, and experience is a reflection"

2. Research approach and framework

This research focuses on the design and analysis of gamified interaction in the VR film creation process, constructing a two-stage level system of construction and deconstruction. It designs a gamified interactive game for the VR film creation process with the core concept of "entering the scene with one shot". The pre-levels realize the gamified transformation of the entire film creation process, from story conception and shot filming; the post-levels reveal the construction mechanism of images through embodied interaction. A generative film agent is added to the game, allowing players to communicate with NPCs and engage in immersive interaction in the game, thereby achieving the de-professionalization of professional film elements. Ultimately, it realizes an interactive closed loop of "creation is experience, experience is level clearing, and level clearing is reflection", enabling users to master the core logic of film creation and cultivate media critical ability in an immersive VR experience, and providing a new practical path for the interactive design and innovative development of VR films.

DOI: 10.54254/2753-7064/2026.BJ33608

3. Methodology

The paper studied mainly about VR interaction design, gamified development, agents and film creation logic. And it uses all kinds of methods comprehensively, like theory study, design thinking, gamification design, embodiment interaction design and so on and iterative prototyping. It is conducive to the research and implementation of design steps one after another to form a cycle, so as to ensure both that the design has sound theory and also takes into account whether it can be actually operated.

3.1. Theoretical research method

In this study,the interactive narrative theory is combined with the immersion theory,generative intelligence technology theory, film audio-visual language theory, media criticism theory. Theoretical analysis refines the immersion and embodiment design principle of VR Media, clarify the logic behind creating a film, summarize the important part of interactive designing with generative AI & VR agent, refer to the concept of image construction and deconstruction in media criticism to provide theoretical support for game levels, interactions, Agent functions.

3.2. Design thinking

The classic design thinking process of "empathy, definition, ideation, prototyping, and testing" is adopted in this study. In the empathy stage, the focus is on the core needs of non-professional users, exploring the immersive experience needs of the general public for film creation and their preferences for gamified interaction. In the definition stage, the core design issues of this game are clarified, namely how to realize the low-threshold gamified transformation of professional film processes, how to improve the interactive experience through VR agents, and how to construct a closed loop of creation, experience, and reflection. In the ideation stage, design schemes are divergently proposed, and the initial conception of the level system, interaction mechanism, and agent roles is completed. In the prototyping and testing stages, the design is implemented and user testing is carried out, and the scheme is optimized according to feedback to ensure that the design meets the actual experience needs of users.

3.3. Gamification design

And then it used the seriousness of games, combined professional film production processes and game features. It transformed all special knowledge about films into game contents, and changed real operational procedures into different levels of steps. The game uses some common game design ideas. They also contain goals establishing, level separation, right now responses and prize arrangements. It breaks down the entire film creation procedure gradually. Its main steps are story planning, shooting arrangement, picture analysis. And it also follows the principle of universality. To take film professions and make them easy game material. Scenes like setting up the shot, cinematic shots etc. So as to make it easier for those without any background in films to learn. Intelligent Agent would also give help at time of need, answer question instantly for users.

3.4. Embodied interaction design

The body features of the VR media are taken as the main foundation by this research. And it makes up the game's interaction based on feeling and space. People can make good links between their self

DOI: 10.54254/2753-7064/2026.BJ33608

and the virtual world by doing physical things. And also it makes all the more immersed. On the level of image construction, it uses many kinds of embodied interaction form. They all have simulated virtual camera movements and such reproduces the practical logic of making a film. In terms of the deconstructed image level, it is embodied design mode. It is mostly depending on the distance sensing and the touch. The users' spatial movement and actual touch of things in the virtual world become main starting points: So as to achieve interactive scenes such like the fade of scenes and characters. It lets people have a kind of perception via physical engagement. People can easily grasp the inner workings of a visual image. It further strengthens the basic idea that people get knowledge through being immersed.

4. Results

Against the backdrop of the digital era with deep integration of AI and VR, this study addresses industry pain points such as the high popularization threshold of traditional filmmaking and fragmented experiences in existing VR film and television interactive products. A VR serious game is designed in which players are accompanied by an AI agent named Lumière throughout the experience, allowing them to learn filmmaking while playing. The bilingual (Chinese-English) interaction website for the film agent is http://47.109.110.156:19999/ (Figure 1).

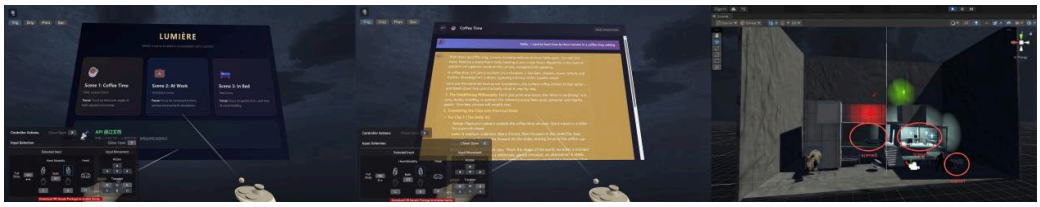

Figure 1. Game agent dialogue and interactive scenes in the theater (picture credit: original)

5. Discussion

5.1. Preparatory levels

5.1.1. Spatial setting of preparatory levels

The spatial setting of this game is a virtual film set space that highly restores the entire film production process. In terms of vision and function, this space is designed with a functionally zoned stage, simulating the working environment of a real film set [6]. Players will enter this space immersively as a new director. At the interaction, the game embeds NPC(Lumière) film intelligent agent dialogue (Figure 2) [7] [8]; the intelligent agent serves as a full-process guide. Players will experience the entire process of film shooting.

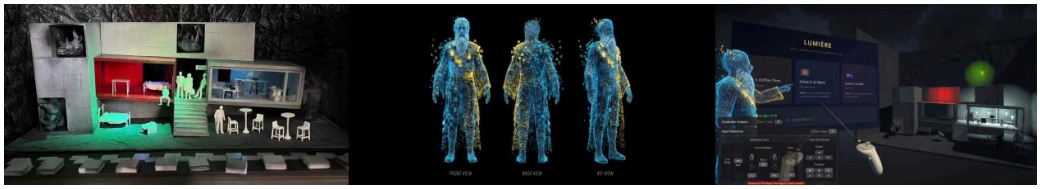

Figure 2. Game space, agent communication interface, three-view diagram and interaction of movie intelligent agents (picture credit: original)

DOI: 10.54254/2753-7064/2026.BJ33608

5.1.2. NPC intelligent agent setting

This game embeds a VR film intelligent agent with generative interaction capabilities. As the core bridge connecting players and the game, the intelligent agent will provide full-process and personalized dialogue services based on the player's operational behaviors and task progress [9]. Accompanied by the intelligent agent, players will sequentially learn creative links such as shot filming. All operational results will be recorded and integrated in real time by the system, and finally generate video clips, realizing the game result from module operation to finished film, allowing players to deeply understand the creative logic of the film industry.

5.1.3. Game completion process

After entering the VR Game, the intelligent agent will communicate with the player [10]. Subsequently, the player can choose three scenes for shooting: the first is a daily leisure scene of drinking coffee, the second is a workplace scene, and the third is a rest scene. The player needs to watch the demonstration video of each scene, then select the correct shot angle (including panorama, close-up, tracking shot, overhead shot, etc.), and complete the answering task. (The flowchart in the figure 3)

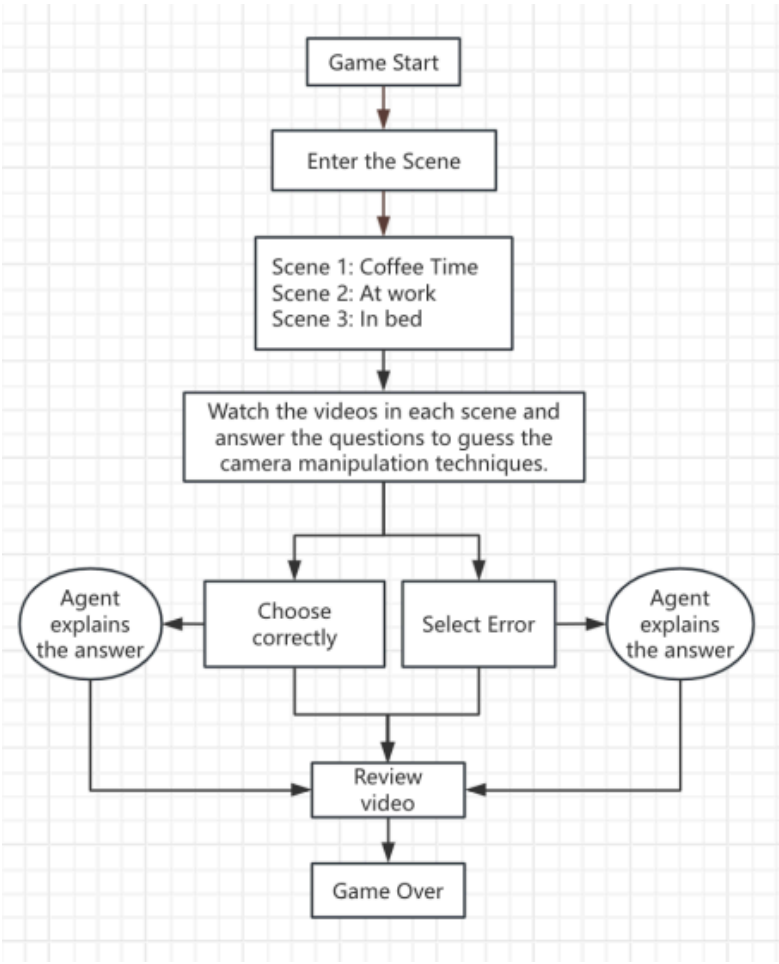

Figure 3. Flowchart (picture credit: original)

DOI: 10.54254/2753-7064/2026.BJ33608

The spatial setting of this game is a virtual film set space that highly restores the entire film production process. In terms of vision and function, this space is designed with a functionally zoned stage, simulating the working environment of a real film set. Players will enter this space immersively as a new director. At the interaction, the game embeds NPC (Lumière) film intelligent agent dialogue; the intelligent agent serves as a full-process guide. Players will experience the entire process of film shooting.

Shot Clip 2 (Core Action): The shot is taken from behind the character's shoulder, showing the character picking up a coffee cup and bringing it close to their mouth with a slight smile on their lips. The background includes a notebook and a desk lamp on the desk. Players need to judge the lens techniques used from four options: "over-the-shoulder shot", "close-up shot", "medium shot", and "subjective shot". The correct answer is "over-the-shoulder shot + close-up shot". After answering correctly, the system will explain the functions, screen effects of the two lens techniques and their fit with the leisure scene; if the answer is incorrect, players can try again.

Shot Clip 3 (Detail Closing): The camera is fixed on the mouth of the coffee cup, capturing the detail that the hot steam weakens after a sip of coffee is taken. Subsequently, the camera quickly shifts to the notebook on the desk. The task is to judge the shooting techniques from four options: "close-up shot ", "whip pan", "panning shot", and "out-of-focus shot". The correct answer is "close-up shot + whip pan". After answering correctly, the explanation can be viewed to understand the detailed functions of the two shooting techniques and the significance of transition foreshadowing.

Scene 2: Working (Workplace Scene, Focusing on Narrative Function, Motion Shots and Atmosphere Building). When the player clicks on Scene 2, they will enter the working scene and complete the answering tasks for 3 shot clips in sequence.

Shot Clip 4 (Scene Establishment): The camera shoots vertically downward from the office ceiling, showing the layout of the entire office area and the working status of employees. Then the camera slowly descends and focuses on the protagonist's workstation by the window with a computer and documents. The task is to judge the shooting techniques from four options: bird's-eye shot, crane shot, long shot, and extreme long shot. The correct answer is bird's-eye shot + crane shot. After answering correctly, the system will explain the functions of the two shooting techniques in scene establishment, perspective switching and foreshadowing subsequent plots. If the answer is incorrect, players can try again.

Shot Clip 5 (Core Working): This shot adopts an eye-level perspective. It centers on the protagonist's facial features and the computer screen. The character presents a slight frown while rapidly operating the keyboard. The protagonist also checks relevant documents intermittently. Subtle camera movements are incorporated throughout the frame to enrich the viewing presentation. The task is to identify the shooting techniques from four options: eye-level angle, close-up shot, handheld shot, and reaction shot. The correct answer is eye-level angle + close-up shot + handheld shot. After answering correctly, the system will explain how the three shooting techniques enhance workplace realism, highlight the character's state, and create a busy rhythm. Players will then proceed to the next clip.

Shot Clip 6 (Emotional Turning Point): The protagonist's facial expression undergoes an obvious emotional shift. It transitions from a state of focused concentration to obvious surprise. This emotional change occurs right after the protagonist receives new information. The camera adopts rapid cross-cutting in this narrative segment. It alternately focuses on the character's face and the computer screen. The message content on the screen is visually blurred for narrative needs. Repeated cuts back to the protagonist's face strengthen the emotional expression. Then the camera pulls back slowly to show the protagonist sitting alone at the workstation with a busy office area in

DOI: 10.54254/2753-7064/2026.BJ33608

the background. The task is to judge the shooting techniques from four options: fast-cut shot, close-up shot, pull-back shot, and reaction shot. The corresponding shooting techniques include fast-cut, close-up, pull-back and reaction shots. These four lens types constitute the standard visual solution for this scene. Upon submitting accurate responses, participants are granted access to systematic theoretical explanations. It further illustrates how these techniques regulate narrative rhythm and intensify character emotions. Meanwhile, they serve as vital visual clues to realize logical narrative connection.

Scene 3: Bedroom Bed (Rest Scene, Focusing on Special Shots, Shot Types and Atmosphere Rendering)When the player clicks on Scene 3, they will enter the bedroom rest scene and complete the answering tasks for 3 shot clips in sequence.

Shot Clip 7 (Opening Foreshadowing): There are no characters in the frame, which presents a full view of the bedroom (half-drawn curtains, sunlight shining on the bed, neatly arranged pillows, and warm water on the bedside table). The camera stays for 3 seconds and then slowly pans toward the bed. The task is to judge the shooting techniques from four options: empty shot, long shot, panning shot, and out-of-focus shot. The correct answer is empty shot + long shot + panning shot. After answering correctly, the system will explain the functions of the three shooting techniques in creating atmosphere, setting up the scene, and guiding attention. If the answer is incorrect, players can try again.

Shot Clip 8 (Core Scene): The camera shoots from a high angle downward, showing the character lying on the bed with eyes closed and breathing evenly, covered with a quilt up to the chest. The picture is slightly hazy with sunlight shining on the character's face, and the camera slowly advances to focus on the character's eyebrows and eyes. The task is to judge the shooting techniques from four options: overhead shot, soft focus shot , push-in shot, and close-up shot. The correct answer is overhead shot + soft focus shot + push-in shot + close-up shot. After answering correctly, the system will explain how the four shooting techniques create a quiet atmosphere, highlight the character's state, and convey a comfortable mood. Players will then proceed to the next clip.

Shot Clip 9 (Closing): The character slowly opens their eyes, and the shot switches to the character's first-person perspective (showing sunlight through the curtain gap and warm water on the bedside table). Then the camera pulls back to show the character lying on the bed and turning to look out the window. Finally, the camera freezes on the sunset outside the window (slow motion). The task is to judge the shooting techniques from four options: subjective shot pull-back shot, slow motion, and medium shot. The correct answer is subjective shot + pull-back shot + slow motion + medium shot. After answering correctly, players can view the explanation to understand the roles of the four shooting techniques in sense of immersion, perspective switching and atmosphere closing. Scene 3 cleared.

After finishing every single one of these three interactive scenes, that smart system would give people an overall summary. This is a summary about the adaptive use of Camera Techniques for different scenes. In order for them to strengthen their own related professional knowledge and actual cognition. Learners can replay the present level and keep on learning. In this way, it can also help them better understand how to shoot cinematically. When the user has completed the learning process they can go on to the next game.

DOI: 10.54254/2753-7064/2026.BJ33608

5.2. Subsequent levels

5.2.1. Embodied interaction mechanism

Level 3: This level's primary arrangement was founded upon what Bailenson said about "When VR Works Well it feels Real to Most People" The technology enhances the immersive and present experience of the VR so there can be true reactions and responses felt in body by the participant for the situation in reality, It makes someone step inside a simulacrum world that was made from behind this medium of media

On which theory base, the research will go into an embodiment interaction after this. Firstly The Players would meet the reality of the VR pictures when they become new Director, then after knowing how picture is formed slowly but surely people start thinking over why people see this specific form of art and finally from being passive immersed in Experience people rise up and contemplate upon Media Structure.

The global interaction mechanism of scene touch degrades which will be used in the following level of this project is constructed. When the player gets to some areas set up for them in the scene, it will change from a real background to showing off all the secret 3D wire frames and green screen grids. When the player is moving back away from the area the scene should return to it's more realistic visuals (figure 4).

The players can naturally change the feeling of complete immersion to think about the creation form of image through this kind of interact virtual-real layout, so as to make people feel and understand the picture more vividly.

Figure 4. Original scene and scene touch degradation interaction (picture credit: original)

5.2.2. Scene interaction mechanism

This project designs an ultimate global interaction mechanism of overall scene collapse. As the player gradually approaches the core of the scene and attempts to touch key props (Figure 5), the system will trigger progressive glitch effects. All objects in the scene gradually become particleized, disintegrate and dissipate completely. The originally complete realistic narrative space eventually degenerates into a blank basic virtual grid environment, where the player stands alone in the digital ruins, completing a closed loop of the experience.

This game worlds forms a three-layer progressive deconstruction logic together with scene touch degradation and character behavior suspension. From local details to the overall space, it dismantles the realistic construction of VR images layer by layer, and finally pulls the player completely out of the immersive narrative illusion, completing the cognitive leap from passive immersion to active media reflection, which deeply echoes the core design concept of this project: from image consumer to image creator.

DOI: 10.54254/2753-7064/2026.BJ33608

Figure 5. Original scene and scene touch degradation interaction (picture credit: original)

6. Conclusion

This study advances interdisciplinary integration and provides implementable design references for the development of VR film and television products and serious games for film and television education. It helps promote the universal popularization of film knowledge and the improvement of public digital media literacy, aligning with the development trend of digital content creation in the AI era. This study builds a complete design system. It actively explores how film can become more immersive and intelligent with AI and VR integration. It transforms ordinary users from passive viewers of film and television content into active creators and rational reflectors, supporting the widespread dissemination and development of film art in the digital era. This research still has certain shortcomings. The VR intelligent module lacks stable functional performance, the presentation of scene decomposition is not smooth enough, and the scale of user tests and personalized adaptive design remain insufficient.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Leininger, P., Weber, C. J., Rothe, S. (2025) Understanding Creative Potential and Use Cases of AI-Generated Environments for Virtual Film Productions: Insights from Industry Professionals//Proceedings of the 2025 ACM International Conference on Interactive Media Experiences. 60-78.

[2] Tian, F., Fu, T.H. and Wu, L.N. (2021) A Comparative Study of Visual Expression between VR and Traditional Cinema from the Perceptual Perspective. Film Art, (5), 136-146.

[3] Zheng, H.H. and Tian, C. (2025) Immersive Narrative Design Based on AIGC: A New Paradigm for Virtual Reality Film and Television Content Production. Project Engineering, 3(3), 50-52.

[4] Li, D.F. (2024) Digital Labyrinths and Waxy Wings: Prospects and Challenges of Generative Intelligent NPCs in Video Games. Journal of Beijing Film Academy, (12), 39-48.

[5] Liu, T. (2026) Generative Montage: An Exploration of Lens Editing with Artificial Intelligence Intervention. Contemporary Cinema, (2), 91-98.

[6] Chen, Y.S. and Cai, R.T. (2025) Application Exploration of Spatial Intelligence Technology in Spatial Mapping of VR Films. Advanced Motion Picture Technology, (2), 32-38.

[7] Guo, J.Q. (2026) From Anthropocene to AIGW: Narrative Autonomy of Generative Images. Film Review, (1), 47-58.

[8] Zhou, W. and Chen, Y.S. (2025) From Non-Player Characters to Artificial Agents: The "Ultimate Life" Narrative of Virtual Reality Media. Editorial Friend, (6), 32-40.

[9] Zhao, L. (2025) From the "Seventh Art" to "Algorithmic Art": Generative AI Reshaping the Landscape of the Film Industry. West China Radio & TV, 46(17), 14-18.

[10] Bailenson, J. N. (2018) Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do.New York: W.W. Norton & Company.