frontiers | Frontiers in Computer Science

TYPE Original Research
PUBLISHED 28 July 2025
DOI 10.3389/fcomp.2025.1591289

OPEN ACCESS

EDITED BY
Shuva Chowdhury,
North Carolina Agricultural and Technical State University, United States

REVIEWED BY
Majid Behravan,
Morgan State University, United States
Abisha D.,
National Engineering College, India

*CORRESPONDENCE
Jie Fu
✉ j.fu1220161@arts.ac.uk

RECEIVED 10 March 2025
ACCEPTED 30 June 2025
PUBLISHED 28 July 2025

CITATION
Fu J, Grierson M, Jiang R, Fu S, He M and Xu M (2025) Interface design and interaction optimization for spatial computing 3D content creation and immersive environment generation using Apple Vision Pro. *Front. Comput. Sci.* 7:1591289. doi: 10.3389/fcomp.2025.1591289

Interface design and interaction optimization for spatial computing 3D content creation and immersive environment generation using Apple Vision Pro

Jie Fu[1]*, Mick Grierson[1], Renpeng Jiang[2], Shun Fu[3], Mengzhi He[1] and Muhan Xu[1]

[1]Creative Computing Institute, University of the Arts London, London, United Kingdom, [2]Dian Jiang Technology Co., LTD, Tianjin, China, [3]Bloks Technology Company, Shanghai, China

Traditional 3D content creation paradigms present significant barriers to meaningful creative expression in XR environments, limiting designers' ability to iterate fluidly between conceptual thinking and spatial implementation. Current tools often disconnect the designer's creative thought process from the immersive context where their work will be experienced, creating a gap between design intention and spatial realization. This disconnect particularly impacts the iterative cycles fundamental to effective design thinking, where creators need to rapidly externalize, test, and refine concepts within their intended spatial context. This research addresses the need for more intuitive, context-aware creation systems that support the iterative nature of creative cognition in immersive environments. We developed Dream Space, a spatial computing system that bridges this gap by enabling designers to think, create, and iterate directly within XR contexts. The system leverages generative AI for rapid prototyping of 3D content and environments, allowing designers to externalize and test creative concepts without breaking their cognitive flow. Through multimodal interaction design utilizing Vision Pro's spatial computing capabilities, creators can manipulate virtual artifacts through natural gestures and gaze, supporting the fluid iteration cycles characteristic of established design thinking frameworks. A mixed-methods evaluation with 20 participants from diverse creative backgrounds demonstrated that spatial computing-based creation paradigms significantly reduce cognitive load in the design process. The system enabled even novice users to complete complex creative tasks within 20-30 minutes, with real-time feedback mechanisms supporting rapid iteration between ideation and implementation. Participants reported enhanced creative flow and reduced technical barriers compared to traditional 3D creation tools. This research contributes to understanding how XR interfaces can better support creative cognition and iterative design processes, offering insights for developing tools that enhance rather than hinder the natural flow of creative thinking in immersive environments.

KEYWORDS

3D content creation, immersive environment generation, Apple Vision Pro, extended reality, multimodal interaction, spatial computing

1 Introduction

1.1 Research background and significance

With the continuous evolution of internet technology and leaps in computational capabilities, we are entering a new era where virtual and real worlds deeply integrate. The metaverse and extended reality (XR) technologies are key to building next-generation internet experiences (Mystakidis, 2022; Dionisio et al., 2013). The metaverse depicts a persistent, shared, and immersive virtual space where users can socialize, entertain, work, and create (Hollensen et al., 2022). XR technologies, including augmented reality (AR), virtual reality (VR), and mixed reality (MR), provide diversified entry points and immersive experiences to the metaverse (Steuer, 1992). To support these visions, efficiently creating and presenting high-quality three-dimensional (3D) content has become a crucial prerequisite (Slater and Sanchez-Vives, 2016).

In metaverse and XR application scenarios, 3D content is not only the basic element for building virtual worlds but also the core carrier of users' immersive experiences. From virtual social spaces and immersive gaming entertainment to remote collaborative offices and virtual education training, to digital twin cities and industrial simulation, various fields require large amounts of high-quality and interactive 3D content (Lee et al., 2022; Bowman and McMahan, 2007). However, traditional three-dimensional content creation methods, such as manual modeling based on modeling software and scanning reconstruction, face numerous challenges in efficiency, cost, and interactivity (Riva and Waterworth, 2001). Manual modeling is time-consuming and labor-intensive, with high professional skill requirements; scanning reconstruction is expensive and difficult to edit and modify (Jerald, 2016). Reality-based modeling techniques have also contributed significantly to surface reconstruction from images (Remondino and Rizzi, 2010). Additionally, 3D content generated by traditional methods often lacks natural and intuitive interaction methods and struggles to fully utilize the immersive characteristics of XR devices (Parisi, 2015).

Particularly noteworthy is that with breakthroughs in generative AI technology, the content creation field is experiencing unprecedented transformation. AI systems based on diffusion models and large language models can now generate complex images, videos, and 3D content from text descriptions (Fitts, 1954). These technologies bring new possibilities for content creation, allowing users to express creative intent in more intuitive and natural ways. Recent progress in hierarchical image generation and multimodal latent spaces, such as the work by Ramesh et al. (2022), further highlights the potential of text-driven immersive content generation. Meanwhile, with the emergence of advanced spatial computing devices like Apple Vision Pro, combining these generative technologies with spatial computing capabilities promises to fundamentally change creation paradigms (Card et al., 1983).

Today's society has growing demands for immersive experiences. Research shows that immersive environments can significantly enhance user engagement and effectiveness in education, training, and creative expression (Sweller, 1988). For example, immersive anatomical learning environments in medical education can improve students' spatial cognitive abilities; immersive prototype displays in product design can more accurately convey design intent; and immersive meditation environments in mental health can provide more effective relaxation experiences (Suchman, 1987). However, the technical barriers to creating these immersive experiences remain high, urgently requiring more convenient and intuitive creation tools (Preece et al., 2015).

Against this background, Dream Space explores a new paradigm of three-dimensional content creation combining generative AI with spatial computing, which has important technological innovation significance and broad application value. This research aims to provide a feasible solution to the above challenges by building an innovative content creation system based on the Vision Pro platform, promoting the popularization of immersive content creation.

1.2 Limitations of traditional three-dimensional content creation

Traditional three-dimensional content creation methods increasingly reveal their limitations when addressing the massive, diversified, and interactive content needs of the metaverse and XR era:

Low Creation Efficiency: Manual modeling is one of the main methods of 3D content creation, but its process is complex and time-consuming. Even experienced modelers may need hours or even days to complete a model of medium complexity (La Viola, 2019), far from meeting the explosive growth demand for metaverse content. A survey of professional 3D modelers shows that creating a commercial-grade character model takes an average of 40-60 hours (Wang et al., 2013), whilst environmental and scene modeling may require weeks or even months. Recent industry surveys also highlight current modeling bottlenecks in professional workflows (Burman and Dubravcsik, 2022). This inefficiency severely constrains the speed and scale of content innovation.

Traditional 3D creation tools mainly rely on traditional 2D input devices such as mouse, keyboard, and touchpad, which are incompatible with the spatial computing and immersive characteristics of XR devices (Makransky and Petersen, 2021). Users often find operations cumbersome and have a steep learning curve when operating these tools in XR environments. Research shows that the learning curve of typical 3D modeling software reaches months or even years (Wu et al., 2022), confining content creation to a small circle of professionals.

Although procedural generation technology has made some progress in specific domains (such as game scene generation), it still has deficiencies in universality and controllability (Gibson, 1979).

Cross-Modal Creation Constraints: Traditional 3D creation tools typically lack effective cross-modal creation support, making it difficult to directly convert text descriptions, sketches, or images to 3D models (Fu et al., 2024). Creators often need to switch between multiple software programmes, from concept design to modeling to texture painting, reducing work efficiency and increasing resistance to creative expression. Especially in the conceptual stage, creators often struggle to materialize ideas in their minds quickly into visible 3D prototypes.

Creation processes typically occur on desktop displays, making it difficult for creators to directly perceive and edit 3D content in immersive environments (Norman, 1988). This limits the intuitiveness and immersiveness of creation and makes it difficult to leverage the advantages of XR devices fully. Research shows that creating in immersive environments can improve spatial understanding ability and creative expression accuracy (43), but traditional tools struggle to provide such experiences.

Collaborative Creation Challenges: As the content scale and complexity increase, the creation process increasingly relies on team collaboration (44). However, traditional tools have limited support for real-time collaborative creation, typically adopting asynchronous workflows, which reduces collaboration efficiency and increases communication costs and version management difficulties. Especially for geographically dispersed teams, the lack of effective remote collaboration tools is a significant constraining factor.

Traditional 3D creation tools are typically complex in design, with professional interfaces that are not user-friendly for ordinary users (45). High software costs, hardware requirements, and steep learning curves make 3D creation the privilege of a few professionals, and it is difficult to achieve large-scale creator participation. With the development of mixed reality, the democratization of content creation has become an inevitable trend, but traditional tools struggle to meet this need.

These limitations collectively constitute the current bottleneck in the 3D content creation field, constraining the development speed and scale of metaverse and XR applications. Therefore, exploring new creation paradigms and combining generative AI with spatial computing technology holds promise for breaking through these limitations and ushering in a new era of immersive content creation.

2 Related work

Current research in 3D content creation mainly focuses on making the process more efficient, improving natural interactions, and creating more immersive experiences. Traditional three-dimensional modeling software such as Blender and Maya, whilst powerful, have steep learning curves and complex operations, making it difficult to meet the demand for rapid content creation in XR environments (Riva and Waterworth, 2001). Scanning reconstruction technology can quickly acquire three-dimensional models of the real world, but equipment costs are high, and subsequent editing and modification are difficult (Jerald, 2016). Procedural generation technology has advanced in certain areas, but still lacks broad applicability and precise control (Gibson, 1979). In particular, rule-based procedural modeling approaches have enabled scalable city generation (Smelik et al., 2010).

To address the limitations of traditional methods, researchers have begun to explore AI-generated, natural-interaction, three-dimensional content creation methods. For example, three-dimensional modeling systems based on gesture recognition and voice control (Garrett, 2010) aim to lower creation barriers and improve efficiency through more natural and intuitive interaction methods.

The rise of spatial computing technology has provided new opportunities for immersive three-dimensional content creation. Collaborative augmented reality systems were among the earliest to explore immersive user interactions (Billinghurst et al., 2002). Spatial computing devices like Apple Vision Pro possess powerful spatial perception and interaction capabilities, laying the foundation for building next-generation three-dimensional content creation tools (Zhai, 1998; Apple, 2023a,b).

In recent years, image-to-3D model reconstruction technology has made significant progress. Methods like Stable Fast, Meshy, and RodinTripoSR (Mann, 2014) can quickly generate high-quality three-dimensional models from single two-dimensional images, providing new approaches for three-dimensional content creation. Platforms like Blockade Lab offer convenient panoramic environment generation APIs, enabling the rapid creation of diverse, immersive environments.

However, spatial computing-based 3D content creation systems are still in the early stages of development, with significant room for improvement in interaction methods, content generation quality, and system performance.

Unlike existing research, this study focuses on the Apple Vision Pro platform, deeply exploring the application of spatial computing technology in three-dimensional content creation and immersive environment generation. This research aims to build an efficient, intuitive, and immersive three-dimensional content creation system, fully utilizing Vision Pro's spatial computing and multimodal interaction capabilities to provide users with a novel creation experience.

3 System architecture design

3.1 Design methodology and foundations

In developing the Dream Space system, we primarily adopted User-Centered Design (UCD) methodology for interface development. We chose this approach because we faced particularly unique challenges, balancing the technical complexity of spatial computing whilst ensuring that ordinary users could intuitively engage with the system.

The design process comprised four distinct phases: firstly, through field observations and in-depth interviews, we identified pain points in users' existing tools, with particular attention to their cognitive patterns and operational habits during 3D creation. Secondly, based on the gathered requirements, we established clear design objectives, discovering that user expectations for spatial interfaces differ significantly from those of traditional 2D interfaces. During the development phase, we adhered to established usability principles while extensively employing rapid prototype testing to validate the effectiveness of each interaction design. Finally, we conducted an objective evaluation of the system through standardized usability scales.

The most crucial finding throughout this process was that users' interaction logic in XR environments differs fundamentally from desktop environments. Traditional hierarchical menu structures cause users to lose orientation in 3D space, leading to the adoption of a spatial position-based functional zoning design. Simultaneously, considering the distinctive nature of multimodal interaction, we established a "primary-auxiliary" interaction paradigm, allowing users to select the most appropriate input method according to the task type, rather than forcing them to use a specific interaction approach.

Our observational methodology, inspired by Suchman's (1987) situated action theory, involved observing authentic user behaviors through testing and interviews, confirming that spatial interaction preferences differ markedly from traditional 2D interfaces. This iterative problem-discovery and problem-solving cycle proved particularly effective in XR interface design, a nascent field where best practices require hands-on exploration.

These design strategies were formulated upon solid theoretical foundations. The design of spatial computing interfaces draws from multiple theoretical frameworks in HCI research. Gibson's (1979) theory of affordances offers fundamental insights into how users perceive and interact with three-dimensional objects, explaining why we found that users' operational preferences in 3D space differ significantly from those in traditional interfaces. Norman's (1988) design principles for everyday things guide us in making complex 3D creation tools intuitive and discoverable, providing the theoretical basis for our development of spatial zoning layouts.

Law Fitts (1954) remains highly relevant for spatial interfaces. However, its application requires careful consideration of depth perception and 3D pointing tasks (Zhai, 1998), which helps us understand why users become disoriented in traditional hierarchical menus. The Model Human Processor framework (Card et al., 1983) helps us understand the cognitive constraints users face when simultaneously processing spatial information and multimodal input, validating our design decision to limit the number of functional modules. Recent advances in spatial cognition research (Montello, 2001; Hegarty, 2004) inform our understanding of how users navigate and manipulate objects in immersive environments. These findings directly influenced our interface design decisions, particularly in relation to spatial memory and wayfinding in virtual spaces.

Throughout the actual design process, each decision was grounded in concrete evidence drawn from both user testing findings and relevant research insights:

Interface Layout:

User testing revealed that when functions were distributed across different positions in 3D space, users frequently struggled to locate control buttons. Drawing upon visual attention research findings, we positioned primary editing functions within users' central field of view, whilst separating visual information from audio feedback to prevent confusion when users process multiple information streams simultaneously.

Gesture Operations:

Initially, we assigned identical functions to all gestures, resulting in frequent user errors. We subsequently recognized that different gesture types should serve distinct purposes: gestures such as pinching proved more suitable for selection actions, whilst pointing gestures better served positioning operations. For complex operations requiring bimanual coordination (such as scaling and rotation), we paid particular attention to rational task distribution between left and right hands, for instance, one hand responsible for grasping whilst the other handles adjustments.

Multimodal Input Coordination:

User feedback indicated that constraining them to single input methods felt unnatural. We discovered that voice input was better suited for conceptual expression (such as "I want a castle"), while gestures proved more effective for spatial operations (moving and rotating), and gaze tracking excelled at object selection. Allowing these input methods to function according to their respective strengths, rather than overlapping or conflicting, resulted in considerably more fluid user interactions.

These adjustments yielded marked improvements in system usability whilst substantially reducing users' learning costs.

3.2 System architecture overview

The system proposed in this research adopts a layered modular architecture, achieving high flexibility and scalability by clearly separating functional components whilst ensuring efficient collaboration between system parts. Through the combination of client, cloud services, and AI services in a three-layer architecture, we've constructed a highly scalable and innovative immersive content creation platform.

We aim to build a spatial computing three-dimensional content creation and immersive environment generation system based on the Apple Vision Pro platform, oriented toward future immersive content production needs. To achieve this goal, this research aims to implement the following specific functions and performance metrics:

In terms of content generation, the system needs diverse cross-modal content generation capabilities: First, the system supports intelligent generation of immersive 360-degree panoramic scene images based on natural language descriptions, providing rich and expandable preset style templates and high-quality texture themes for users to choose, quickly replace, and fine-tune, allowing user groups including professional designers to efficiently build virtual environments with diverse styles and controllable quality; To meet the higher requirements of professional creation processes for asset reuse and flexibility, the system supports users in uploading custom 360-degree panoramic background images and integrating them into immersive creation environments, achieving deep customization of scene atmosphere and visual style. Second, the system needs to implement intelligent and efficient conversion from two-dimensional images to high-quality three-dimensional models, and continuously explore the possibility of directly generating complex three-dimensional models based on text and voice instructions, especially for common model assets in virtual scenes, the system supports users in uploading custom three-dimensional models, currently only compatible with USDZ format, with plans to support mainstream three-dimensional formats in the future, and should provide convenient model asset management and reuse mechanisms to enhance professional creation efficiency;

For common elements in virtual scenes, such as ground/floors, the system supports users in fine-tuning pattern customization, editing, and hiding to meet the needs of scene details and brand personalisation design. All generated and imported three-dimensional content must support professional-grade, immersive visualization presentation on the Vision Pro platform, ensuring that creation results can fully meet the stringent standards of professional XR content production for visual quality and immersive experience.

For human-computer interaction, this research focuses on the innovative application of multimodal natural interaction methods: The system will fully exploit Apple Vision Pro's spatial computing

potential, deeply integrating multimodal sensor inputs such as gesture recognition and gaze tracking to build a natural, intuitive, and efficient human-computer interaction mechanism. Users should be able to directly manipulate three-dimensional models through natural gestures in immersive environments, achieving precise rotation, scaling, translation, and other operations. Meanwhile, combined with gaze tracking technology, the system should support gaze-based scene editing and object selection, achieving more fine-grained and intent-driven content editing and scene construction, minimizing user learning costs and enhancing the efficiency of the creation process.

In terms of system performance and user experience, this research pursues performance and smooth user experience: The system achieves practical operation levels in three-dimensional model generation quality, immersive environment rendering effects, and interaction responsiveness, ensuring that generated 3D models are rich in detail while interaction operations are smooth, natural, and responsive. The ultimate goal is to provide users with a smooth, efficient, immersive, and enjoyable three-dimensional content creation tool, allowing them to freely unleash creativity on the Vision Pro platform and efficiently produce high-quality XR content. The overall architecture of the Dream Space system is illustrated in Figure 1.

3.3 Client layer

The client layer is the frontend interface for user interaction with the system, developed on the Apple Vision Pro platform. This layer fully utilizes Vision Pro's spatial computing capabilities, providing an intuitive creation environment by integrating frameworks such as SwiftUI, ARKit, RealityKit, and Metal. The client interface adopts a zoned management design concept divided into three parts: scene management area, toolbar area, and editing area. The scene management area at the top displays the current scene name and timestamp and provides reset and save functions, allowing users to save creation progress or restore default states at any time. The vertical toolbar area on the left contains quick-switching buttons for three main functional modules: 3D models, virtual space (360° environment), and ground settings, allowing users to directly click to enter the corresponding editing mode. The central editing area dynamically changes according to the currently selected function module, displaying corresponding control options.

We used SwiftUI to build these spatialised interface elements, implementing toolbars and interaction panels that conform to spatial computing characteristics. Interface elements dynamically adjust the layout according to the user's gaze and spatial position, maintaining optimal visual effects and interaction distance. For example, when users approach to view a 3D model, related editing controls appropriately enlarge and move to the center of the field of view; when users step back to observe the overall scene, these controls automatically shrink and move to edge positions, avoiding obstructing vision.

The 3D model editing interface provides precise transformation controls, including adjustment in three dimensions: Translate, Rotate, and Scale. Each dimension is equipped with intuitive slider controls. Users can make independent precise adjustments along the X, Y, and Z axes or use the Reset button to quickly restore the model's default state. Additionally, the system provides duplicate and delete functions, allowing users to quickly create multiple similar models or remove unwanted content. The virtual space editing interface allows users to select different types of environment representations, including sphere image, sphere video, SkyBox, and SkyDome. Users can choose to display or hide the virtual space through the "hide" switch for better focus on model editing. The system provides default texture and custom import options and supports adding new environment maps through "New Import" or "AI Generation" buttons.

The ground settings interface provides ground type selection (such as Flat plane) and texture mapping control. Users can enable or disable texture mapping and bump mapping functions to enhance the visual effect and realism of the ground. Similarly, users can control the display state of the ground through the hide switch.

RealityKit serves as the core rendering engine responsible for the real-time presentation of high-quality three-dimensional scenes. It supports high-fidelity rendering of complex models and simulates natural lighting and physical effects, providing users with an immersive visual experience. By leveraging Vision Pro's built-in gesture recognition system, we precisely bind users' hand movements to functions in the application, achieving a natural and smooth operation experience. This approach, processed directly on the device, eliminates network latency, allowing users to control and edit virtual content as intuitively as manipulating real objects, greatly enhancing the immersion and efficiency of the creation process. Users can also conveniently organize and access their creation content through local data management functions, forming a complete creation workflow. Based on detailed feedback from the first round of user testing, we restructured the user experience framework, optimizing interface layout and interaction flow and laying a more user-friendly foundation for subsequent testing.

3.4 AI service layer

The AI service layer is the intelligent core of the system, integrating multiple cutting-edge AI technologies to build a complete process from content understanding to generation. The design focus of this layer is to provide high-quality intelligent generation capabilities whilst ensuring the controllability and efficiency of the generation process. The core AI generation module fuses multiple generation technologies, forming an intelligent creation pipeline that works collaboratively. TripoSR technology is responsible for converting two-dimensional images into three-dimensional models, with the advantage of being able to extract rich geometric and semantic information from a single image, quickly reconstructing structurally complete 3D models. The TripoSR architecture used here is based on recent profile-guided mesh reconstruction advances (Yariv et al., 2023). We've made multiple optimizations to the TripoSR algorithm, particularly in feature extraction and topology inference, enabling it to better handle complex shapes and texture details. Interestingly, user feedback indicates that TripoSR performs

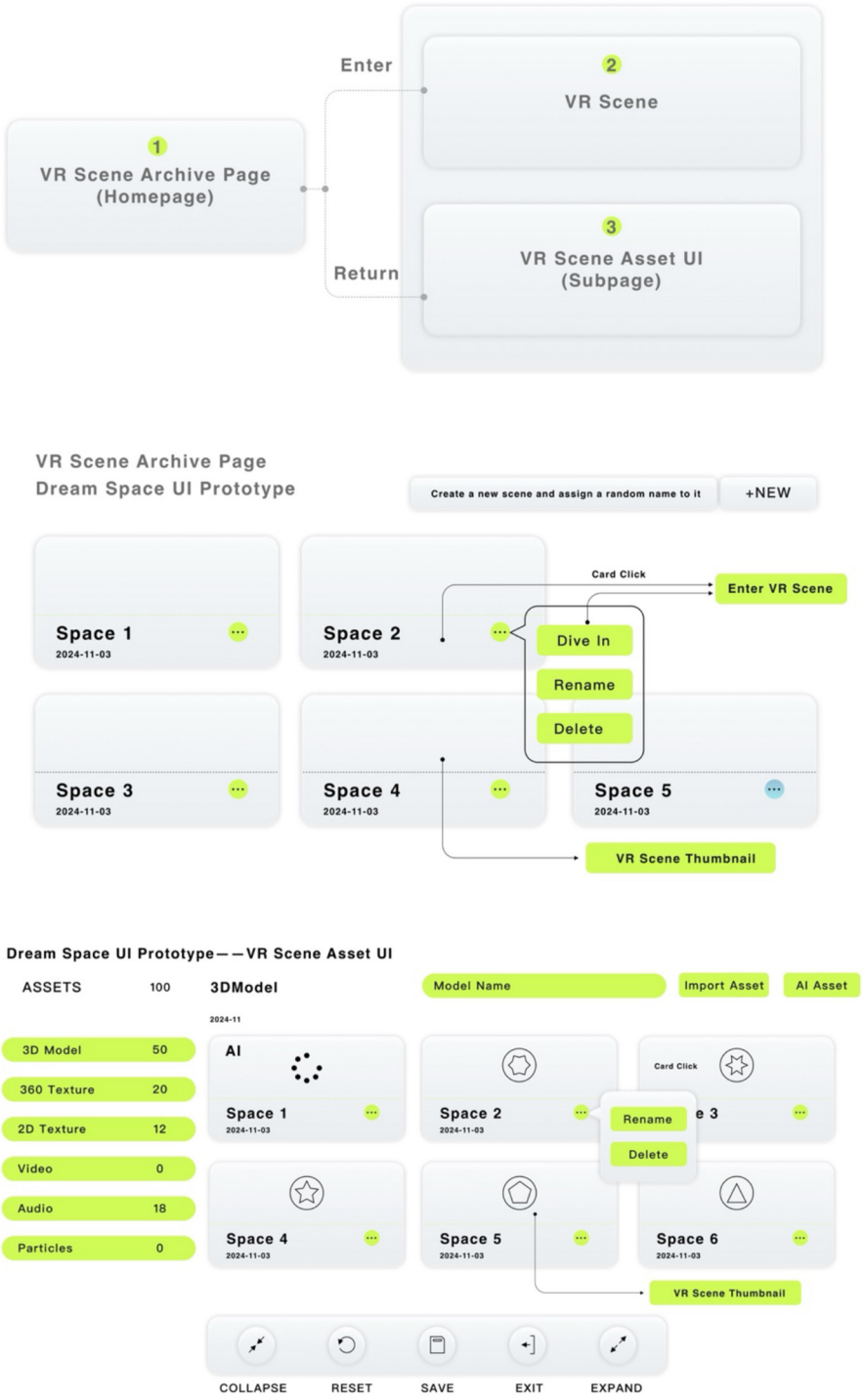

FIGURE 1
Dream space system architecture diagram.

better in text generation mode than in image generation mode, possibly because text descriptions provide more explicit semantic information and structural features, reducing visual ambiguities and allowing the algorithm to construct three-dimensional expressions more in line with user expectations. At the same time, text descriptions are not limited by a single perspective, providing more comprehensive object concept information and helping the algorithm infer complete three-dimensional structures.

Blockade Lab integration provides the system with high-quality environment generation capabilities. Through AI technology based on diffusion models, the system can convert simple text descriptions into stereoscopic panoramic environment textures, supporting users to quickly create diverse and immersive scenes.

This module not only supports stylised artistic scenes but can also generate photo-realistic natural environments, meeting different creation needs.

To ensure the quality of generated content, during three-dimensional reconstruction and model optimization, these algorithms can intelligently identify and preserve the generation of immersive scenes, such as the sky and edges of 360-degree generation, sharp edges, surface continuity, and texture details, ensuring that original visual features are maintained even after model simplification. The virtual space interface supports quick switching between multiple environment representation forms, allowing users to select between image sphere, video sphere, SkyBox, or SkyDome via radio buttons. The texture preview area uses a grid layout, allowing users to intuitively compare different environment effects. "New Import" and "AI" buttons are located in the upper right corner of the texture area, facilitating users to expand their environment resource library at any time.

This layered modular design improves the maintainability and scalability of the system and lays the foundation for future integration of more generation technologies and interaction methods. With the rapid development of AI technology, the system can continuously enhance performance and functionality by updating specific modules without restructuring the entire architecture.

4 System implementation

The three-dimensional content creation system implemented on the Vision Pro platform in this research contains three core components: three-dimensional model generation, immersive environment generation, and intelligent interaction control. These components work together to form a complete creation ecosystem. We developed the interface and runtime architecture based on visionOS capabilities (Apple, 2023b).

Throughout the implementation process, our technical choices were grounded in practical considerations. Through user observation, we discovered that users' creative processes follow distinct stages, initially selecting the environmental atmosphere, then placing three-dimensional models, and finally adjusting scene details such as ground elements. Based on this finding, we structured the system into corresponding functional modules, ensuring that users' operational pathways align with the system architecture, thereby reducing cognitive load and making interactions feel more intuitive. Regarding interaction design, user testing revealed that different operational tasks are best suited to different input methods. Eye gaze proves most natural for object selection, gestures feel most intuitive for moving and rotating objects, whilst voice input works most effectively for expressing creative concepts. Consequently, rather than enforcing uniform interaction methods, we allowed each input modality to leverage its particular strengths, enabling users to select the most suitable operational method for specific tasks freely. For the AI integration strategy, we positioned AI to handle initial content generation while maintaining users' control over outcomes. Users can modify generated models or environments at any time, completely replace them, or regenerate content entirely. This approach leverages AI's generative capabilities and mitigates the creative limitations that may arise from complete reliance on AI. The entire system's technical implementation centers on enhancing users' creative abilities rather than replacing their creative judgement.

4.1 Immersive environment generation

The three-dimensional model generation component allows users to convert two-dimensional images into interactive three-dimensional models. The system first preprocesses input images, including noise reduction and color adjustment, to improve image quality. Subsequently, we use optimized neural networks to extract geometric and semantic features from images. The core three-dimensional reconstruction is based on an improved version of the TripoSR algorithm, which, under our optimization, can better maintain the topological structure and detailed features of objects.

After optimization, the generated models are loaded into RealityKit scenes for user interaction. The process is designed with intuitive progress indicators, allowing users to understand the processing status. Through a series of algorithm optimizations, we've shortened the image-to-model conversion time from the original 125-130 seconds to 80-102 seconds, greatly enhancing creation fluidity.

4.2 Intelligent interaction control

Interaction methods designed for traditional desktop environments are not suitable for spatial computing platforms. We've developed an interaction framework integrating gesture recognition, gaze tracking, and voice input, allowing users to interact with three-dimensional content in a more natural way. On the Vision Pro platform, we've embedded gesture-tracking functionality into the Dream Space application through custom code, fully utilizing the platform's native capabilities. Users can directly grab, rotate, and scale models with gestures, select operation objects through gaze, or execute complex operations by combining voice commands.

The system can fuse signals from different input channels, recognize user intent, and execute corresponding operations. For instance, users can adjust the size of a model with gestures whilst gazing at it, or activate specific editing functions through voice commands. We've also implemented predictive caching mechanisms, where the system predicts possible next steps based on current operations and preloads required resources, further reducing interaction latency.

This multimodal interaction method improves operation efficiency and significantly lowers the learning threshold. In our user testing, even participants encountering the system for the first time could master basic operations within 5 minutes and complete full creation tasks within 20 minutes.

4.3 Interface design and interaction optimization

In Vision Pro's spatial computing environment, we've rethought interface design principles to create an intuitive yet

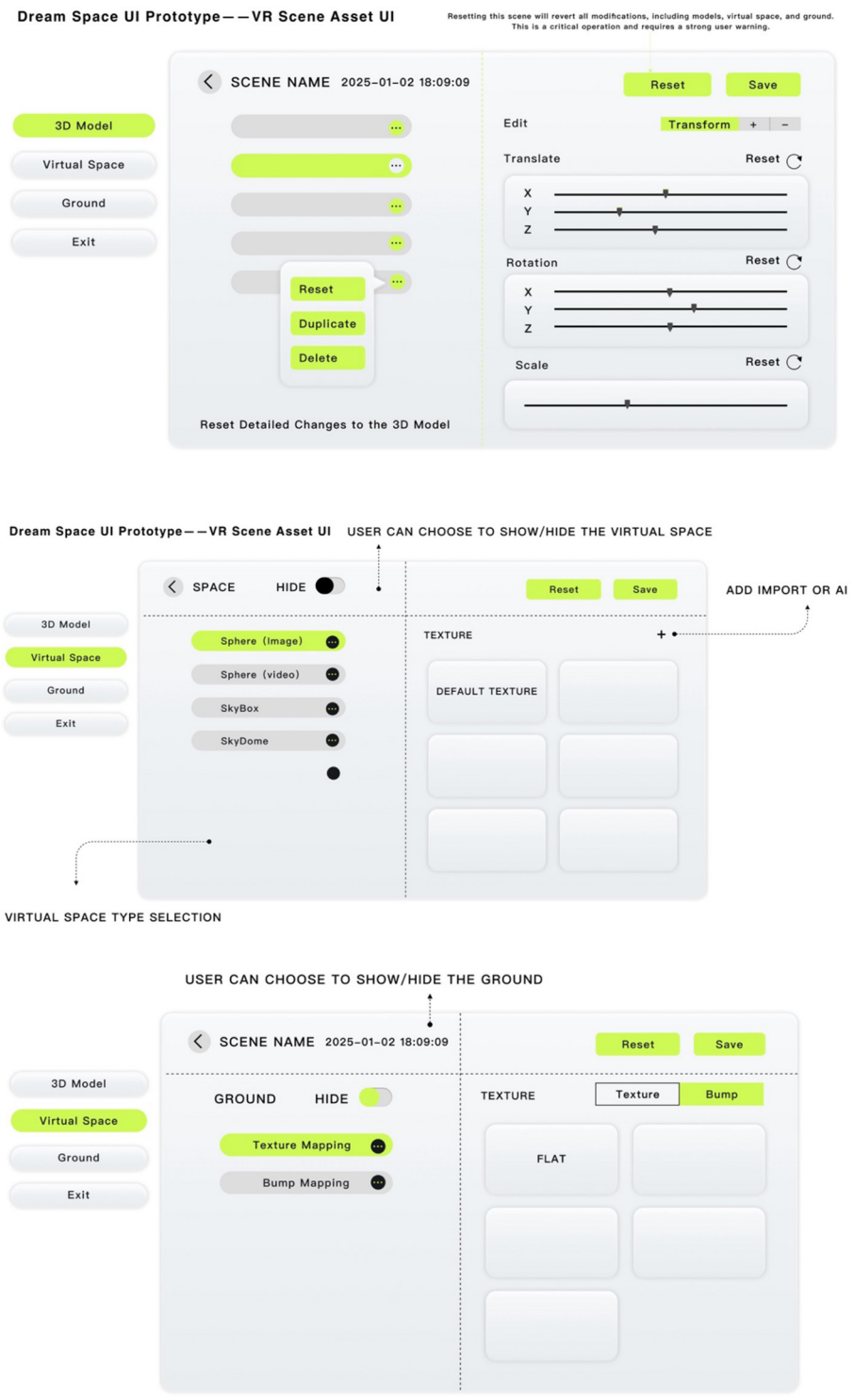

FIGURE 2
Dream Space System Architecture Diagram (Second Iteration).

efficient interaction experience. As shown in the Figure 2, the system interface is divided into three major functional areas: 3D model editing, virtual space (360°) management, and ground settings. This zoned design allows users to quickly locate needed functions, reducing operation paths.

We pay special attention to safety prompt mechanisms for dangerous operations. For example, when users attempt to reset a scene, the system displays clear warning prompts: "This operation will reset all object changes, including models, virtual space, and ground," effectively preventing unintended loss of creation results

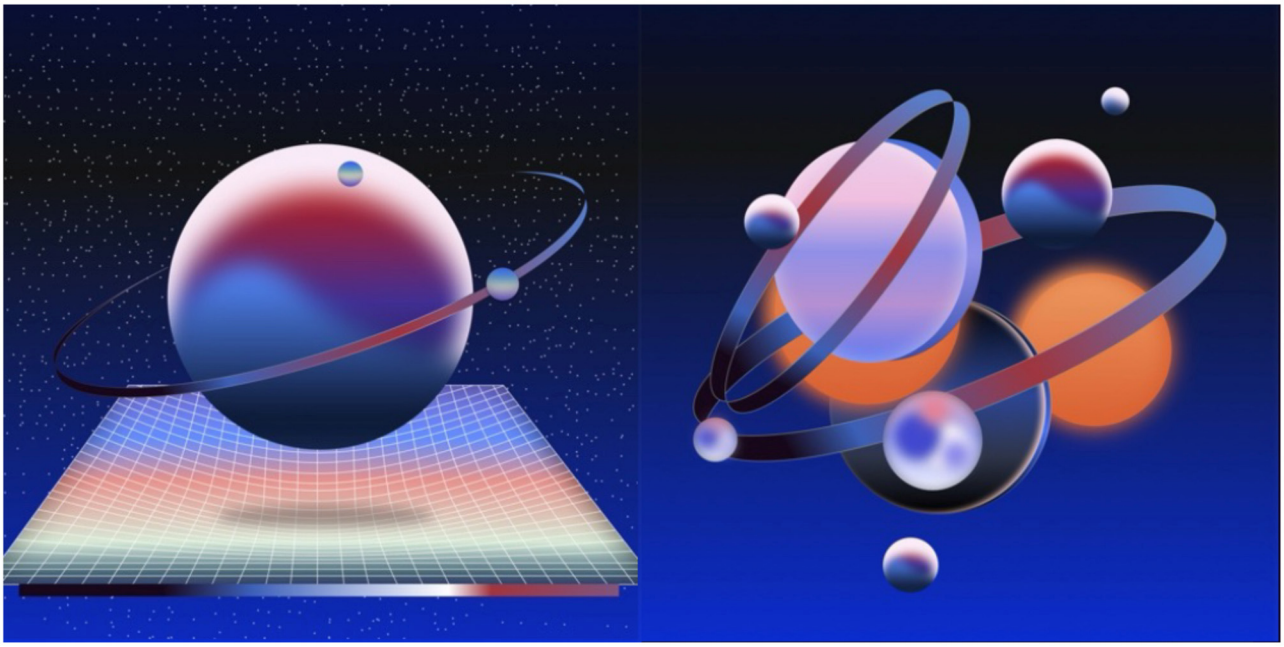

FIGURE 3
Dream Space XR Icon design.

due to misoperation. This preventive design ensures users can make clear decisions before executing operations that might affect work progress. The 3D model editing interface adopts a combined control scheme, supporting both precise numerical adjustment and direct spatial operation. Translation, rotation, and scaling control panels are all equipped with bidirectional sliders, allowing users to precisely adjust values for each dimension. Meanwhile, cyclic arrow buttons to the right of each control group allow users to quickly reset individual transformation properties without affecting other settings.

Through the Vision Pro platform's spatial interaction capabilities, users can directly manipulate models using natural gestures whilst seeing real-time updated, precise values. This interaction method greatly enhances editing intuitiveness and efficiency. This multi-level control design demonstrates excellent efficiency in user testing, particularly suitable for professional creation scenarios requiring precise adjustments, whilst also providing a friendly operation experience for beginners. We've fully considered operational characteristics in spatial computing environments, developing a "gaze+gesture" combined interaction mode. For example, when users gaze at a model and make specific gestures, the system recognizes user intent and executes corresponding operations. This interaction method, which doesn't require explicit "clicking," better aligns with spatial computing's natural interaction philosophy, greatly lowering the learning threshold.

Based on early user testing feedback, we found users often encountered "drift" issues when performing precise control in space, especially when fine-tuning model positions. To address this issue, we developed an adaptive precision control system that automatically reduces sensitivity when detecting users performing fine adjustments, improving operation precision whilst maintaining higher sensitivity during large-range movements, ensuring smooth operation. Interface layout and control dimensions were also carefully designed and tested through multiple rounds. Main function buttons are set to a virtual size of not less than 2cm, ensuring users can accurately touch them; important operations (such as save, reset) use prominent color coding, with the save button using orange and the reset button using blue to enhance visual recognition; secondary functions use gray tones to reduce visual interference. The visual identity of Dream Space XR is depicted in Figure 3.

Based on feedback needs from different user groups, the system provides optimized interface configurations. The first draft simplified the interface according to feedback and was updated with more refined model control options. Figure 4 presents the complete functional layout of the Dream Space user interface, including 3D model editing, virtual space selection, and ground controls. These updates make the system easy to use and capable of meeting professional creation needs. Interface design and interaction optimization measures collectively form an intuitive, efficient creation environment, allowing users to focus attention on creative expression rather than the tool operation itself. Figure 5 shows the early-stage user interface tested in both simulator and real user environments. User testing results show that even participants using spatial computing devices for the first time can master basic operation processes in a short time and begin creation activities. Figure 6 provides a screenshot of immersive environment generation from the first system version.

4.4 Performance optimization

To achieve a smooth creation experience on spatial computing devices like Vision Pro, we implemented comprehensive

FIGURE 4
Dream Space User Interface Design: Complete Functional Module Layout for 3D Model Editing, Virtual Space Management, and Ground Settings.

performance optimization strategies. These optimizations focus on rendering performance and cover memory management, generation processes, and interaction response, collectively forming a multi-level performance assurance system. Figure 7 illustrates the system's performance during 3D content generation tasks.

4.4.1 Model complexity management

We developed a complete three-dimensional model optimization framework, with adaptive mesh simplification technology at its core, capable of controlling model complexity whilst maintaining visual quality. Mesh simplification methods like those proposed by Fuhrmann et al. (2003) provide useful

10.3389/fcomp.2025.1591289

FIGURE 5
The initial UI interaction design of Dream Space XR, as demonstrated in both the simulator and real user environments.

FIGURE 6
Screenshot of Environment Generation Testing in the First Version of Dream Space XR.

FIGURE 7
Screenshot of 3D Content Generation Testing in the First Version of Dream Space XR.

precedents for maintaining detail under polygon constraints. This framework contains several key links:

Model Type Recognition and Differentiated Processing: The system applies different simplification strategies based on model characteristics. For organic models (such as characters, animals), priority is given to preserving contours and key feature points; for hard surface models (such as architecture, machinery), emphasis is placed on maintaining edge sharpness and surface flatness.

Optimization Workflow: We integrated core functionalities of multiple professional tools, including ZRemesher topology reconstruction to convert irregular triangle meshes into regular quadrilateral meshes, Catmull-Clark subdivision algorithms to improve surface smoothness, and Decimation Master to control polygon count. Precise face count control: Through experimental validation, we determined the optimal face count control range: ordinary interaction models controlled at 80,000 faces, focus models requiring close observation maintained at around 100,000

faces, and for lightweight needs in real-time rendering, further reduction to about 10,000 faces using Simplygon.

Texture Optimization and Detail Preservation: Using UV Master to automatically generate and optimize UV layouts and baking high-model details into normal, displacement, and ambient occlusion maps to ensure visual quality is not compromised.

From Single Processing to Batch Optimization: We have completed optimization cases for complex organic models such as coral, successfully reducing polygon count from 1.5 million to 52,000. The next step is implementing batch optimization through custom algorithms and automated programming, establishing a batch processing system capable of simultaneously handling multiple models (Fu et al., 2024). AI-generated coral structures and their visual variations are shown in Figure 8.

4.4.2 Material and texture optimization

In material optimization, we adopted a multi-level strategy, closely integrated with TripoSR's texture processing technology. First, the system automatically applies compression algorithms to uploaded and TripoSR-generated textures, converting most textures to ASTC format, which can reduce texture memory consumption by 60-75% whilst maintaining visual quality. Second, we implemented automatic mipmap generation and optimization, ensuring high texture sampling efficiency when observed from a distance and reducing visual artifacts such as moiré patterns. Proper handling of projection distortion is also essential for model realism (Blinn, 1992).

TripoSR can already provide excellent initial textures when generating models, but to further optimize performance, we developed a specialized post-processing workflow. For complex materials generated by TripoSR, the system performs intelligent simplification. For example, merging multi-layer PBR materials into single-layer materials or baking procedurally generated textures into static maps. Especially for high-detail surfaces generated by TripoSR, we use normal maps and ambient occlusion maps to preserve details whilst significantly reducing geometric complexity. Figure 9 demonstrates the second version of Dream Space with support for stylisation and asset imports.

4.4.3 Interaction design methodology

The interface design of Dream Space was grounded in established interaction design frameworks, particularly Human-Centered Design (HCD) as outlined by Norman (2013) and ISO 9241-210. Our approach followed an iterative design cycle involving low-fidelity prototyping, internal heuristic evaluations, and user-centered self-assessment to identify usability challenges early in development. The use of think-aloud protocols, semi-structured interviews, and task-based observation informed critical design choices—such as gesture–gaze combinations, tool panel placements, and adaptive control precision—that directly addressed real user needs.

This process aligned with design thinking principles and the Double Diamond framework, enabling divergent exploration of ideas followed by convergent refinement based on user evaluation. Each iteration incorporated direct user feedback to reduce cognitive load, enhance spatial intuitiveness, and promote creative flow.

By embedding these methods into the development pipeline, Dream Space evolved from a technical prototype into an experience-oriented immersive tool, offering accessible, intuitive interaction for both novice and professional creators. This methodological grounding bridges the gap between system functionality and user-centered interface design, providing a solid theoretical foundation for our implementation and evaluation strategy.

5 User research and evaluation

To comprehensively evaluate the practicality and user experience of our system, we conducted systematic user research across two distinct phases of testing. This research aimed to collect behavioral data and subjective feedback from users, providing a robust foundation for subsequent optimization.

We employed two rounds of user testing to assess the system's usability and user experience. We chose this approach primarily because XR interface design remains relatively nascent, with many issues only emerging during actual use, purely theoretical analysis or heuristic evaluation might overlook critical usability problems that become apparent in practice.

The first round of testing employed primarily qualitative methods, inviting users from diverse backgrounds, including researchers, designers, and general users. We encouraged users to think aloud during their interactions, enabling us to understand their genuine feelings and points of confusion. However, within XR environments, we discovered that traditional think-aloud protocols required adjustment; users found it difficult to verbalize whilst performing spatial operations, so we incorporated post-task reflection sessions where users could review their experiences after completing tasks.

The second round introduced quantitative metrics, including task completion times and operational error rates, whilst employing the NASA-TLX scale to evaluate users' cognitive load (Hart and Staveland, 1988). We also utilized the UEQ-S to measure overall user experience satisfaction. These standardized instruments helped us assess differences more objectively before and after system improvements, particularly regarding changes in user burden and satisfaction levels. The UEQ-S has been validated as a concise user experience instrument (Schrepp et al., 2017). Examples of user-generated 3D scenes are shown in Figure 10.

The combination of both testing rounds enabled us to gain deep insights into users' specific problems and requirements, whilst providing data to validate the effectiveness of our improvements.

5.1 Research user research and evaluation

We adopted a two-phase mixed research method involving different assessment approaches. Participants across both phases spanned ages from 18 to 50 with varied professional backgrounds including computer science, machine learning, musical composition, botanical research, marketing and AR/VR development, ensuring representativeness across our findings. All participants provided informed consent prior to testing.

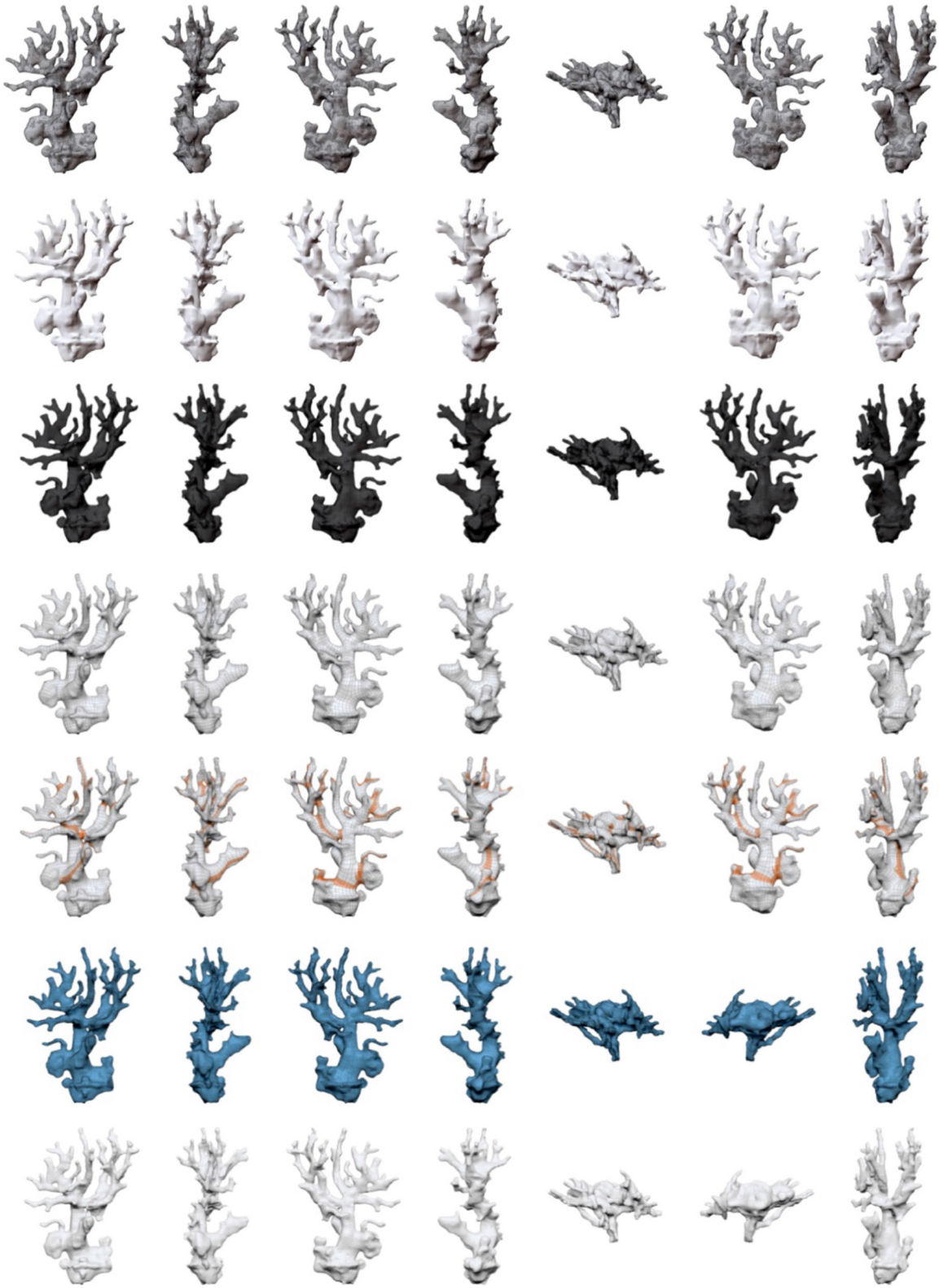

FIGURE 8
AI-Generated 3D Model and Optimization Process (Fu et al., 2024).

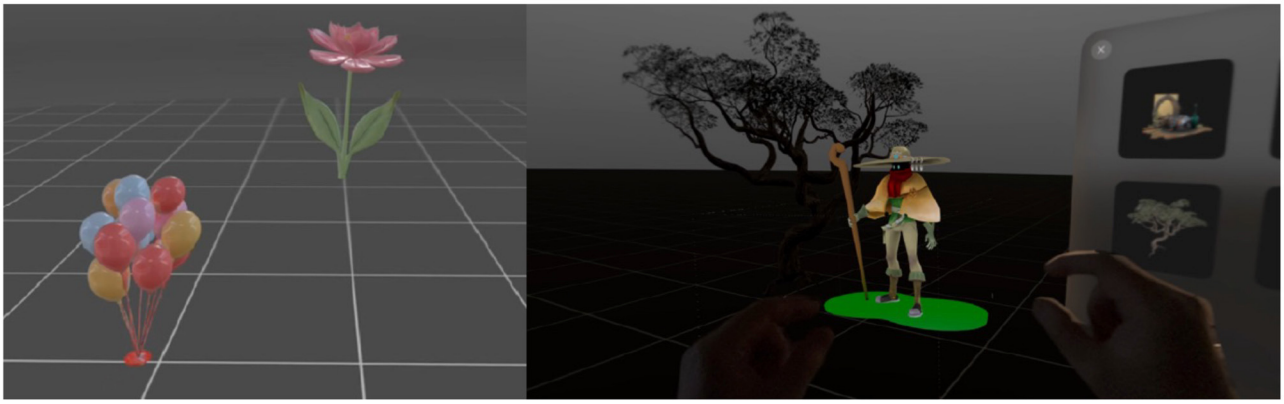

FIGURE 9
The second version of Dream Space XR introduces model stylisation and supports importing assets from a model library.

5.1.1 First phase: qualitative exploratory research

The first phase primarily employed qualitative research methods, inviting 10 participants from diverse backgrounds including human-computer interaction researchers, three-dimensional designers and ordinary users. Researchers used a think-aloud approach, recording participants' operational behaviors and gathering feedback as they completed preset tasks, supplemented by semi-structured interviews following the testing sessions.

We designed three core tasks to evaluate the system: uploading images and text and converting them to 3D models, generating and adjusting immersive environments, and controlling 3D models through gestures. This phase emphasized depth of understanding over quantitative metrics, focusing on user perceptions, difficulties encountered, and creative possibilities identified. Figure 7 illustrates the system's performance during 3D content generation tasks.

5.1.2 Second phase: mixed-method validation research

The second phase employed a mixed-method approach, selecting 10 different participants, each with testing sessions lasting approximately one hour. This phase incorporated both qualitative feedback and quantitative evaluation methods to record task completion time, operation error rate and subjective satisfaction ratings using a 7-point Likert scale.

This approach allowed us to validate improvements implemented after the first testing phase whilst collecting systematic measurements of performance enhancements. The combination of rich qualitative insights with quantitative metrics provided a comprehensive evaluation framework for system assessment.

5.2 First round user feedback analysis

The initial testing phase yielded valuable qualitative insights regarding the system's strengths and limitations:

5.2.1 Immersive experience and environment feedback

Users particularly emphasized the advantage of "VR immersive experience superior to 2D experience" and expressed appreciation for "360-generated scenes and styles," validating the correctness of our design direction in spatial computing environments. One participant with an art background pointed out "VR experience is better," indicating that we need to further enhance immersion in augmented reality scenarios.

Users suggested multiple application scenarios, including "museum displays," "VR meditation spaces (with music)," "virtual learning spaces," and "game scene design," expanding the application prospects of the system. One participant with a technical background believed the system "can be used for decorating DIY spaces," further confirming its potential for personal creative expression.

5.2.2 Interface and control challenges

Significant usability issues were identified in the first round. Most prominently, users reported difficulties with the interface design, noting that the control panel was obstructive and requesting options to hide UI elements to enhance immersion. Many users also expressed difficulties with precise object manipulation, with insufficient visual feedback when selecting items.

Professional creators emphasized the importance of precise control, proposing specific suggestions such as adding auxiliary line functionality (similar to Gravity Sketch), implementing Snap functionality for quick return to initial values, and adding single-axis rotation capabilities. One interaction designer suggested using "text as buttons" to simplify the interface and reduce interaction layers.

5.2.3 AI generation functionality feedback

First-round participants provided constructive opinions on AI generation functions. One creative worker suggested, "Can we select different styles in AI 3D (realistic vs abstract) like 360

FIGURE 10
Partial User-Generated Testing Screenshots from the Second Version of Dream Space XR.

scene generation?" and pointed out issues such as "too few floor generation options." Another participant proposed the specific application of "adding AI generation of relatively abstract art functionality to coral."

A professional 3D modeler pointed out, "image generation speed needs optimization" whilst affirming "native transformer effects are good" and "3D model generation effects are good," providing a basis for our algorithm selection. One designer suggested "adding a function to view historical prompts," reflecting users' need for traceability in the generation process.

5.2.4 Technical function and innovation suggestions

A professional developer proposed multiple technical optimization suggestions, including "improving texture generation and light and shadow effects," "adding reset buttons," and "frame needs to be hideable." They also suggested supporting "scene programming" and "video content (especially 360 videos)," indicating higher expectations for system function expansion.

A researcher emphasized the necessity of "supporting uploading users' own models" and suggested "providing templates

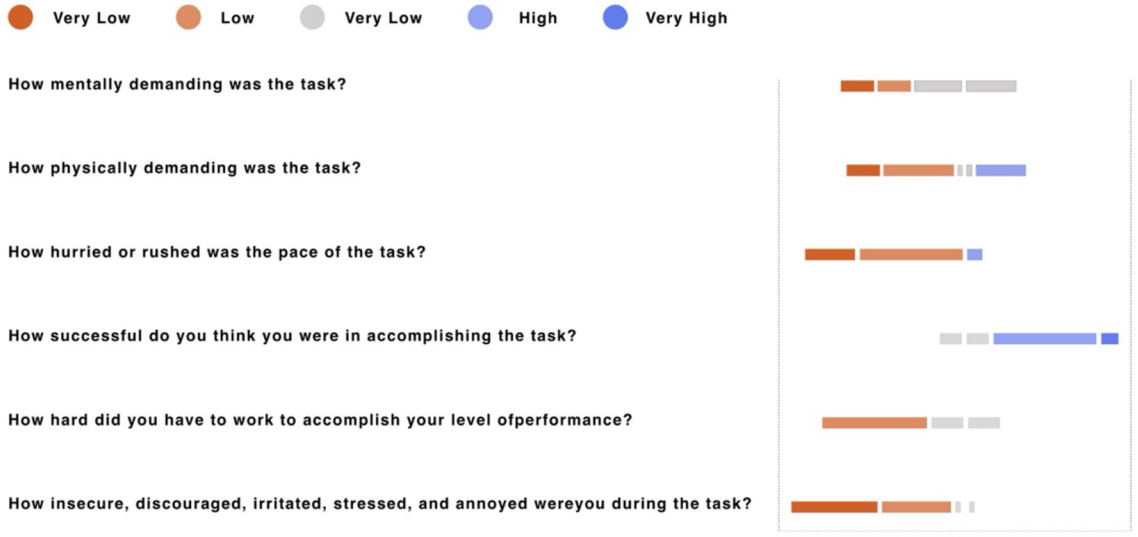

FIGURE 11
User ratings of task demands, performance, and emotional response while using the software, measured on a five-point Likert scale from Very Low to Very High.

for quick testing such as 'underwater,' 'city,' 'forest,' etc." to attract new users' attention. These suggestions pointed toward improving work efficiency and professional applications.

Based on the first round of user testing, we implemented several improvements before the second testing phase, including enhanced object selection indicators, optimized AI generation buttons, voice input control for three-dimensional model generation, and improvements to the object deletion mechanism.

5.3 User experience and task load assessment

The user evaluation (n=10) revealed positive results across both experience and usability metrics. The UEQ-S questionnaire showed the system received positive ratings in both Hedonic Quality (1.75) and Pragmatic Quality (1.00), with an Overall Score of 1.38 (scale:−3 to +3). Notably, the system performed exceptionally well in creating an "exciting" experience (1.90) while maintaining acceptable usability standards. The NASA-TLX assessment complemented these findings, with approximately 80% of users reporting "high" to "very high" task success rates and predominantly "low" to "medium" mental demands and stress levels. All participants (100%) indicated they would use the software in the future. Figure 11 shows users' cognitive load evaluations across six NASA-TLX dimensions. The system's stronger performance in emotional aspects compared to pragmatic features suggests development priorities should focus on improving usability (scored 0.50) while maintaining the engaging qualities that users valued. Though these results are promising, a larger sample of 20-30 participants would provide more definitive conclusions in future evaluations. Detailed scores for user experience and workload metrics are presented in Figure 12.

The Dream Space system received positive user experience ratings (average UEQ-S score of 1.4) whilst maintaining a relatively low cognitive workload (average NASA-TLX score of 2.0), demonstrating that the system delivers both a pleasant user experience and operates without imposing excessive operational pressure on users. Figure 13 compares UEQ-S and NASA-TLX distributions via box plots.

5.4 Second round user feedback analysis

The second testing phase, combining qualitative feedback with quantitative measurements, revealed both improvements from our initial adjustments and areas requiring further development:

5.4.1 Quantitative performance improvements

Our measurements revealed that after system optimization between testing phases, users significantly reduced their completion time for image-to-3D model and environment generation tasks. Although the time required for model gesture operation tasks still needs improvement, the operation error rate decreased substantially following the Video introduction, as participants gained a better understanding of the gesture tracking mechanisms. This indicates a significant enhancement in interaction precision across testing phases.User satisfaction questionnaires showed significant improvements in ratings for operation convenience, interface friendliness, functional practicality, and immersive experience. These data validate the effectiveness of our iterative design approach and confirm the system's improved usability after initial modifications.

5.4.2 System strengths and creative potential

Users in the second round consistently praised the system's immersive quality and creative capabilities, particularly when generating visually engaging fantasy environments. The VR interface received positive feedback for its intuitive navigation and

METRIC / SCALES	MEAN	SD
UEQ-S: Pragmatic Quality	1.00	0.88
UEQ-S: Hedonic Quality	1.75	1.06
UEO-S:Overall	1.38	0.90
NASA-TLX: Mental (LIB)	2.40	0.84
NASA-TLX: Physical (LIB)	2.50	1.18
NASA-TLX: Time (LIB)	1.90	0.88
NASA-TLX: Performance	3.80	0.63
NASA-TLX: Load (LIB)	2.40	0.52
NASA-TLX: Frustration (LIB)	1.60	0.70
NASA-TLX: Overall (LIB)	2.00	0.49

FIGURE 12
Descriptive statistics (Mean and Standard Deviation) for user experience and workload metrics. UEQ-S measures pragmatic and hedonic quality, while NASA-TLX assesses cognitive and physical workload across various dimensions.

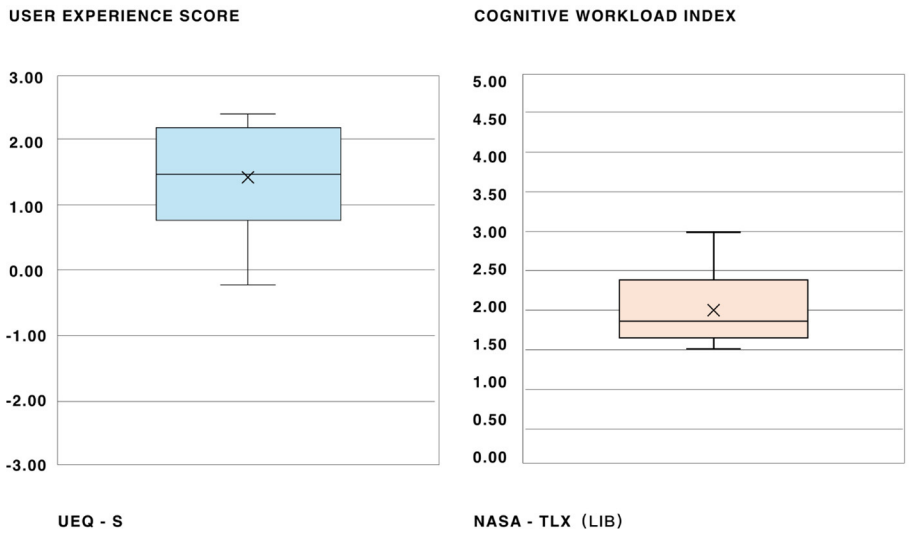

FIGURE 13
Box plot comparison of UEQ-S and NASA-TLX scores. The UEQ-S score represents user experience quality, while the NASA-TLX score reflects perceived workload. The plots illustrate the distribution, median, and variability of responses.

object manipulation functionality, which several participants noted as significantly more efficient than traditional 3D workflows. One participant with a computer science background emphasized: "In VR, the sense of immersion is much stronger, and the ability to move objects intuitively makes iterative adjustments much more efficient than on a computer."

The system's ability to rapidly translate conceptual ideas into explorable virtual spaces emerged as a particularly valued feature. A participant with a machine learning background mentioned how the system effectively "visualized scenes from science fiction novels" they had read, creating a novel form of engagement with imaginative content. This capacity for translating mental imagery into three-dimensional spaces was cited as having significant potential for educational applications, creative design processes and scientific visualization. Participants' willingness to adopt the software in the future is summarized in Figure 14.

5.4.3 Persistent technical challenges

Despite improvements, users still identified several technical challenges. System stability remained an issue, with two reported crashes during resource-intensive operations in testing sessions. Our performance tests indicate that in extreme scenarios (over 30 high-detail models with a total face count exceeding 2 million),

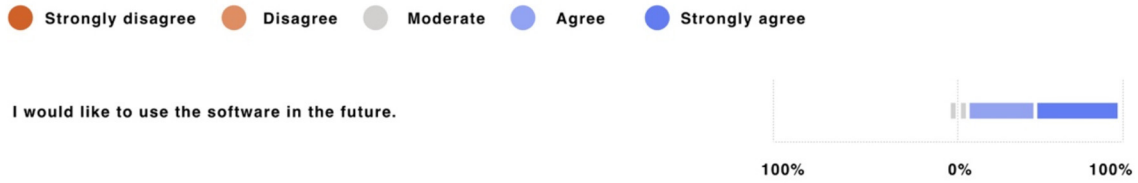

FIGURE 14
User agreement on future use of the software, rated on a five-point Likert scale from Strongly Disagree to Strongly Agree.

frame rates drop to around 40fps, which, although still acceptable, is not the optimal experience.

A significant theme across feedback was the perceived disconnect between generated environments and placed 3D objects. Several users noted that models appeared somewhat "fixed" or "rigid" when positioned within scenes, lacking stylistic and lighting coherence with their surroundings. A participant with computer science expertise suggested that "models could be optimized to better match the environment" to create more convincing integrated scenes.

The precision and style control of model generation needs further optimization in specific scenarios, particularly when generating models with complex structures or fine details, where the system sometimes cannot fully capture detail features, especially for organic models such as characters or animals.

5.4.4 Expanded application potential

Second-round participants expanded on potential applications, with particularly noteworthy suggestions about enhancing emotional engagement with data by presenting information within contextually relevant environments—for instance, displaying endangered species statistics within natural habitat backgrounds.

A participant working in academic research suggested the system could create "immersive VR knowledge graphs" where concepts are presented interactively in three-dimensional space, potentially transforming how literature reviews and conceptual relationships are visualized. Others highlighted potential applications in architectural visualization, scientific data exploration, and therapeutic environments.

Several participants proposed suggestions for building social functions, including "adding sharing functionality," "establishing a public gallery," and "supporting user co-creation," demonstrating expectations for a creative community environment. A participant with an industrial design background focused on cross-media applications, suggesting "supporting the generation of 3D printable models" and "adding souvenir generation functionality," emphasizing the importance of connections between virtual and physical reality.

5.5 User feedback-driven system iteration planning

Based on in-depth findings from both rounds of user research, we formulated an iteration roadmap for system optimization. This planning fully considers the difficulty of technical implementation and the priority of user needs, ensuring the system can develop continuously and robustly whilst maintaining sensitivity and responsiveness to user feedback.

5.5.1 Short-term priorities

The primary task is to address issues encountered by users in basic interaction and content generation, improving system usability and efficiency. We plan to optimize the visibility of AI generation buttons, improve the sensitivity of three-dimensional rotation, and add key functions for model management, such as one-click deletion, recovery, and copying.

User demands for floor effects will also be met, including height adjustment, hiding options, and richer material expressions. In terms of generated content, voice input support will complement gesture operations, particularly suitable for quickly triggering functions or inputting generation prompts. We will introduce style selection functionality, allowing users to freely switch between realistic and abstract styles, better controlling creative intent.

The most pressing concern is improving the integration between objects and environments, creating more cohesive visual scenes where models seamlessly blend with their surroundings through matched lighting, stylistic consistency and appropriate scaling. This aesthetic integration will be complemented by enhanced dynamic capabilities, allowing for subtle animations and environmental effects that bring static scenes to life.

Interface requires refinement to balance comprehensive functionality with immersive experience; we plan to implement collapsible control panels and more intuitive selection feedback whilst maintaining easy access to creative tools. Performance optimization remains crucial, with system stability during complex operations needing significant attention to prevent crashes.

Based on the first round of user feedback, we have already implemented two important improvements: adding voice input control for three-dimensional model generation and significantly optimizing the object deletion mechanism, making operations more intuitive and efficient.

5.5.2 Mid-term planning

Mid-term planning will focus on enhancing the system's interaction capabilities and creative possibilities. Historical record functionality will help users track and reuse successful creative experiences, whilst environmental light customization options will further enhance scene atmospheric expressiveness. We will also explore the feasibility of adding particle effects and lighting functions in space, providing creators with richer visual expression tools.

Content generation will be enhanced through expanded model libraries with greater stylistic diversity, improved texture quality at higher magnifications, and more sophisticated prompt guidance to help users articulate their creative visions. In direct response to user feedback, our development plans include a comprehensive expansion of the 3D asset library with categorized models for users to select from, addressing the frequent requests for greater variety and stylistic options. Additionally, a dynamic animation system will be implemented, enabling both objects and characters to be animated within the environment, fulfilling users' desires for more lively and interactive scenes.

5.5.3 Long-term vision

In the long term, social functionality will become an important direction for system development. By adding scene screenshots and sharing functions, users can more easily record and disseminate their creative results. Public model libraries and remix functions will promote resource sharing and creative inspiration within the community, forming a positive and creative ecosystem.

Another possibility is exploring transformation paths from virtual to physical, supporting three-dimensional printing, and allowing digital creations to cross the boundaries between virtual and reality, providing users with a complete creative expression experience.

Multi-person collaboration functionality is another important future direction, but network latency and data synchronization challenges must be addressed. In spatial computing environments, real-time sharing of three-dimensional content and interaction states requires efficient network transmission and state synchronization mechanisms, which still face challenges under current mobile device bandwidth and computational capabilities.

6 Discussion and future work

We have successfully built a spatial computing three-dimensional content creation system based on the Vision Pro platform, exploring the potential of this technology in immersive content creation. Research shows that our system has made progress in multiple aspects.

At the technical level, the system integrates spatial computing frameworks and optimizes model generation algorithms and interaction methods, effectively improving performance and user experience. The zoned management interface and multimodal interaction controls we developed based on Vision Pro's spatial interaction characteristics significantly lowered the learning threshold and improved operation precision, particularly in user interface design. The system's modular design gives it scalability and maintainability, allowing flexible adaptation to future technological changes and user need evolution. User research confirms that spatial computing-based creation methods have gained user recognition, surpassing traditional tools in efficiency and satisfaction.

Compared to traditional creation tools, Dream Space improved task completion efficiency by over 50%, with user satisfaction also significantly increased. These data fully validate the enormous potential of spatial computing technology in the three-dimensional content creation field.

6.1 Integration prospects of generative AI and spatial computing

The rapid development in the generative AI field brings revolutionary opportunities for spatial computing content creation. Recent "one-image-to-3D-world" technology demonstrated by World Labs and the Genie 2 foundation world model released by DeepMind has already shown powerful capabilities to generate complete 3D scenes from single images or text descriptions. These breakthrough advances will profoundly change creation methods and efficiency. Interface usability remains a key challenge, particularly for novice creators (Hammon, 2021).

Compared to single object generation technologies, these new world models can understand and generate complete three-dimensional environments, including terrain, architecture, vegetation, and atmosphere. When these technologies integrate with XR devices, creators will be able to generate complete immersive environments in minutes through simple descriptions or sketches, whilst traditional methods might require weeks or even months.

Our Dream Space system has already laid the foundation for this integration. The system's modular design allows us to flexibly integrate new AI models as technology develops, whilst our developed spatial interaction framework provides users with intuitive capabilities to manipulate and edit this generated content. In the future, we plan to integrate advanced world models like Genie 2 into the system, further enhancing the scale and quality of content generation.

Particularly noteworthy is that content generated by these world models possesses inherent physical consistency and semantic understanding capabilities, which will make interactions in XR environments more natural and intuitive. For example, generated environments may already include correct collision areas and physical properties, allowing users to directly physically interact with objects in the environment without needing additional settings and adjustments.

6.2 Future research directions

Based on current research results and technological trends, we've planned four main future research directions:

6.2.1 Professional domain application expansion

We will apply the system to professional domains such as education, design, architecture, and healthcare, exploring the value of spatial computing and AI in professional scenarios. For example, in architectural design, precise three-dimensional model dimension control can be achieved through AI like Text to CAD, quickly generating and experiencing architectural spaces, evaluating lighting, circulation, and spatial perception; in cultural heritage protection, precious artifacts and historical scenes can be reconstructed and virtually displayed. Spatial understanding is crucial in immersive architectural planning and simulation (Li et al., 2020).

Application scenarios suggested in user research, such as museum displays, aquarium scenes, and virtual learning spaces,

will become our key exploration directions. These professional applications not only validate the system's practical value but will also drive system functionality in more professional directions.

6.2.2 Multimodal interaction innovation

We will explore more natural, intuitive multimodal interaction methods, further lowering spatial creation barriers. Research directions include combining gaze and voice interaction to provide a more direct expression of intent and developing context-aware interaction modes where the system can automatically adjust interaction methods and interface layouts based on user behavior patterns and environmental characteristics.

We plan to explore the fusion possibilities of brain-computer interface technology with existing gesture recognition systems, creating a truly intuitive "thought control" experience. The application of haptic feedback in spatial interaction is also a key research direction. By providing physical sensations through wearable devices, users will be able to obtain the texture and feedback of virtual objects, significantly enhancing the certainty and precision of operations.

These tactile elements will complete spatial editing experiences, providing necessary physical boundary sensations for fine operations. Notably, the full implementation of these advanced interaction technologies may need to wait for further optimization of MR hardware. We position these innovations as expandable directions when technology matures, outlining a future within reach for users where creation processes will be more natural and fluid, infinitely approaching physical operations in the real world.

6.2.3 AI generation model optimization

With the rapid development of generative AI technology, we will continuously improve AI models integrated into the system, enhancing the quality, diversity, and controllability of generated content. Specific directions include developing more precise style control mechanisms allowing users to adjust visual styles while maintaining content consistency; improving model understanding of professional domain knowledge, generating structurally more accurate and functionally more reasonable professional content; optimizing resource utilization efficiency of generation processes, and reducing processing time and energy consumption. Morphing-based mesh transformations may further support flexible structure generation (Yuan et al., 2020).

Particularly worth noting is the integration of world models like Genie 2, which will fundamentally enhance the system's generation capabilities, allowing users to create more complex, coherent virtual environments. We will research how to maintain the generation capabilities of these models whilst providing suitable editing and customization tools, ensuring creative freedom.

We have successfully built a spatial computing three-dimensional content creation system based on the Vision Pro platform, exploring the potential of this technology in immersive content creation. Research shows that our system has made progress in multiple aspects.

At the technical level, the system integrates spatial computing frameworks and optimizes model generation algorithms and interaction methods, effectively improving performance and user experience. The zoned management interface and multimodal interaction controls we developed based on Vision Pro's spatial interaction characteristics significantly lowered the learning threshold and improved operation precision, particularly in user interface design. The system's modular design gives it scalability and maintainability, allowing flexible adaptation to future technological changes and user need evolution. User research confirms that spatial computing-based creation methods have gained user recognition, surpassing traditional tools in efficiency and satisfaction.

Compared to traditional creation tools, Dream Space improved task completion efficiency by over 50%, with user satisfaction also significantly increased. These data fully validate the enormous potential of spatial computing technology in the three-dimensional content creation field.

The iterative improvements made throughout the interface development process strongly reflect the principles of Human-Centered Design. Feedback from both testing phases directly informed key design refinements, including spatial layout adjustments, interaction simplification, and rebalancing of the visual hierarchy. These changes contributed to minimizing cognitive load and enhancing usability, particularly for novice users.

This alignment between theory and practice demonstrates the practical value of HCD and design thinking frameworks in guiding the development of immersive interfaces. In the context of XR, where spatial cognition and multimodal interaction are still emerging challenges, grounding design in user needs and iterative evaluation becomes essential for meaningful engagement.

6.2.4 Cross-platform collaboration ecosystem construction

In the future, we plan to build a more open, collaborative, social cross-platform creation ecosystem, achieving seamless collaboration and content sharing between users of different devices and platforms. This includes developing lightweight Web and mobile versions, allowing non-XR device users to participate in creation processes; establishing cloud-based content libraries and sharing platforms, supporting the discovery and reuse of creative resources; designing real-time collaboration mechanisms, and allowing multiple users to simultaneously create and communicate in the same virtual space.

Community building is also an important direction; we will establish points and reputation systems, encouraging high-quality content creation and sharing; develop content review and copyright protection mechanisms, ensuring healthy community development; explore cooperation with educational and creative institutions, promoting system applications in broader domains.

7 Conclusion

The spatial computing three-dimensional content creation system Dream Space proposed in this research provides an efficient, intuitive, and innovative solution for immersive content creation. By integrating advanced AI generation technology with spatial computing capabilities, we've achieved intelligent conversion from two-dimensional images to high-quality three-dimensional models and immersive environment generation based on natural language.

The system's multimodal interaction framework, combining gesture recognition, gaze tracking, and voice input, provides users with an unprecedented natural creation experience.

Despite the identified challenges, participants demonstrated significant enthusiasm for the system's potential. As one user with a PhD in music summarized: "It feels like translating imagination into three-dimensional space in a way that's intuitive and immediately explorable." This sentiment encapsulates the system's core value proposition whilst acknowledging the remaining technical hurdles that must be addressed.

The system's modular design ensures high performance and provides possibilities for future integration of more advanced AI models and interaction technologies. With breakthrough advances from institutions like World Labs and DeepMind in the generative AI field, our system architecture is ready to seamlessly integrate these emerging capabilities, further expanding creation possibilities.

From a long-term perspective, our research results will help promote the democratization process of immersive content creation. By lowering creation barriers, improving efficiency, and enhancing expressiveness, Dream Space XR makes high-quality XR content creation no longer limited to professional teams but open to a broader creator community. This transformation will bring richer and more diverse content to the metaverse and XR applications, promoting the popularization and development of these emerging platforms.

At the intersection of technology and creativity, Dream Space XR is not just a creation tool but a window looking toward the future, giving us a glimpse of how spatial computing technology will change human expression and creation methods. With continuous advancement and refinement of technology, spatial computing-based three-dimensional content creation systems will play an increasingly important role in future metaverse and XR ecosystems, bringing revolutionary changes to the construction and experience of the digital world.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

JF: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft. MG: Supervision, Writing – original draft. RJ: Software, Writing – original draft. SF: Resources, Writing – original draft. MH: Formal analysis, Writing – original draft. MX: Formal analysis, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The article processing fee charge was supported by the University of the Arts London (UAL).

Acknowledgments

The authors would like to express sincere gratitude to development partner Jiang Renpeng for his invaluable contribution to the system implementation. Special thanks to Meng Zhihe and Muhan Xu for assistance during the testing phase and Shun Fu for help with three-dimensional aspects of the project. We are deeply grateful to Professor Mick Grierson for his insightful guidance and steadfast support throughout this research. We also extend our appreciation to all volunteers who participated in user testing, whose feedback has been instrumental in refining the system.

Conflict of interest

RJ was employed at Dian Jiang Technology Co., LTD. SF was employed at Bloks Technology Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI Statement

The author(s) declare that Gen AI was used in the creation of this manuscript. As the author, I confirm and take full responsibility for the use of generative AI in the preparation of this manuscript. In our APP development, we have employed various AI-related plugins: immersive scene generation primarily based on Blockade Labs technology, whilst Text to 3D and Image to 3D functionalities utilize Tripo SR technology. It is worth noting that we have made customized adjustments to these foundational technologies to meet our specific Dream space XR's Vision Pro application requirements. Moreover, the user generation testing phase has also fully utilized integrated versions of these technologies, ensuring consistency and completeness of the user experience. In subsequent development, we may replace the relevant APIs.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2025.1591289/full#supplementary-material

References

Apple (2023a). *Apple vision pro*. Available online at: https://www.apple.com/apple-vision-pro/ (Accessed October 26, 2023).

Apple (2023b). *visionOS - technology - apple developer*. Available online at: https://developer.apple.com/visionos/technology/ (Accessed October 26, 2023).

Billinghurst, M., Weghorst, S., Furness, T. A., and Shared, I. I. I. (2002). Space augmented reality. *Sci. Am*. 287, 84–91.

Blinn, J. F. (1992). Models of distortion. *ACM Siggraph Comp. Graph*. 26, 177–178. doi: 10.1145/142413.996911

Bowman, D. A., and McMahan, R. P. (2007). Virtual reality: how much immersion is enough? *Computer* 40, 36–41. doi: 10.1109/MC.2007.257

Burman, J., and Dubravcsik, A. (2022). "The state of 3D modeling in industry: a survey of professionals," in *ACM SIGGRAPH 2022 Talks* (New York, NY: ACM), 1–2.

Card, S. K., Moran, T. P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dionisio, J. D., Gilbert, R., and Choi, G. C. (2013). 3D virtual worlds and distance learning. *Comput. Educ.* 52, 75–93. doi: 10.1145/2480741.2480751

Fitts, P. M. (1954). The information capacity of the human motor system. *J. Exp. Psychol.* 47, 381–391. doi: 10.1037/h0055392

Fu, J., Fu, S., and Grierson, M. (2024). Coral model generation from single images for virtual reality applications. *arXiv*. doi: 10.48550/arXiv.2409.02376

Fuhrmann, A., Goesele, M., Langguth, F., et al. (2003). Automatic 3d model simplification. *Comp. Graph Forum* 22, 37–55. doi: 10.1111/1467-8659.00651

Garrett, J. J. (2010). *The Elements of User Experience: User-Centered Design for the Web and Beyond*. Berkeley, CA: New Riders.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Hammon, E. (2021). Building 3D models with professional approaches versus tools aimed at novices. *Int. J. Art. Des. Educ.* 40, 330–344. doi: 10.1111/jade.12357

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Hegarty, M. A. (2004). dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 175–191. doi: 10.1016/j.intell.2003.12.001

Hollensen, S., Kotler, P., and Opresnik, D. (2022). Metaverse - the new reality? *J. Bus. Strategy* 43, 349–357. doi: 10.1108/JBS-02-2022-0036

Jerald, J. (2016). *The VR Book: Human-Centered Design for Virtual Reality*. San Rafael, CA: Morgan & Claypool. doi: 10.1145/2792790

La Viola, J. J., Jr. (2019). Natural interaction. *Proc. IEEE*. 107, 1061–1064. doi: 10.1109/JPROC.2019.2907277

Lee, L. H., Braud, T., Zhou, P., et al. (2022). All one needs to know about metaverse: a complete survey on technological singularity, virtual ecosystem, and research agenda. *J. Internet Serv. Appl.* 13:37. doi: 10.1186/s13174-022-00151-z

Li, W., Agrawala, M., Curless, B., and Salesin, D. (2020). Spatial understanding in architectural design analysis. *ACM Trans. Graph.* 39, 1–14. doi: 10.1145/3414685.3417763

Makransky, G., and Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. *Educ. Psychol. Rev.* 33, 1–22. doi: 10.1007/s10648-020-09586-2

Mann, S. (2014). "Wearable computing as means for personal empowerment," in *2014 International Symposium on Technology and Society (ISTAS): social Implications of Wearable Computing and Ubiquitous Sensors* (Piscataway, NJ: IEEE Press), 1–8. doi: 10.1109/ISTAS.2013.6613094

Montello, D. R. (2001). "Spatial cognition and environmental behavior," in *Handbook of Environmental Psychology* (Hoboken, NJ: Wiley), 318–341.

Mystakidis, S. (2022). Metaverse. *Encyclopedia* 2, 486–493. doi: 10.3390/encyclopedia2010031

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York, NY: Basic Books.

Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. New York, NY: Basic Books.

Parisi, T. (2015). *Learning Virtual Reality: Developing Immersive Experiences and Applications for Desktop, Web, and Mobile*. Sebastopol, CA: O'Reilly Media.

Preece, J., Sharp, H., and Rogers, Y. (2015). *Interaction Design: Beyond Human-Computer Interaction*. Chichester: John Wiley & Sons.

Ramesh, A., Dhariwal, P., Nichol, A., et al. (2022). Hierarchical text-conditional image generation with CLIP latents. *arXiv*. doi: 10.48550/arXiv.2204.06125

Remondino, F., and Rizzi, A. (2010). Reality-based 3D modeling and surface reconstruction from images. *Photogramm Rec.* 25, 269–292. doi: 10.1111/j.1477-9730.2010.00599.x

Riva, G., and Waterworth, J. A. (2001). Presence in mediated environments. *Cyberpsychol. Behav.* 4, 477–496. doi: 10.1089/109493101750527033

Schrepp, M., Hinderks, A., and Thomaschewski, J. (2017). Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). *Int. J. Interact. Multimedia Artif. Intell.* 4, 103–108. doi: 10.9781/ijimai.2017.09.001

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI*. 3:74. doi: 10.3389/frobt.2016.00074

Smelik, R. M., Tutenel, T., de Kraker, K. J., de Jong, P., and Verbree, E. (2010). Rule-based procedural modelling of virtual city models. *Comp. Graph.* 34, 345–355. doi: 10.1016/j.cag.2010.03.006

Steuer, J. (1992). Defining virtual reality: dimensions determining telepresence. *J. Commun.* 42, 73–93. doi: 10.1111/j.1460-2466.1992.tb00812.x

Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving. *Cognit. Sci.* 12, 257–285. doi: 10.1207/s15516709cog1202_4

Wang, X., Love, P. E. D., Kim, M. J., Park, C. S., Sing, C. P., Hou, L. A., et al. (2013). conceptual framework for integrating building information modeling with augmented reality. *Autom Constr.* 34, 37–44. doi: 10.1016/j.autcon.2012.10.012

Wu, H., Wang, J., Zhang, X., et al. (2022). A comparative study of online and offline 3D design platforms: an educational perspective. *Comput Educ.* 180:104425. doi: 10.1016/j.compedu.2022.104425

Yariv, R., Gu, J., Kulkarni, T., Tulsiani, S., Srivastava, N., Barron, J. T., et al. (2023). "TripoSR: direct profile-guided single-image 3D mesh reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (New York, NY: IEEE), 6527–6536.

Yuan, Y., Lai, Y. K., Yang, J., et al. (2020). Data-driven shape interpolation and morphing editing. *Comp. Graph Forum*. 39, 697–710. doi: 10.1111/cgf.14121

Zhai, S. (1998). User performance in relation to 3D input device design. *ACM Siggraph Comp. Graphics* 32, 50–54. doi: 10.1145/307710.307728