A comparison between two XR-based tabletop solutions for 3D storyboarding

Virtual Reality (2026) 30:144
https://doi.org/10.1007/s10055-026-01381-8

ORIGINAL ARTICLE

A comparison between two XR-based tabletop solutions for 3D storyboarding

Federico Manuri[1] · Francesco De Pace[2] · Valentina Carrescia[1] · Sara Piumatti[1] · Andrea Sanna[1]

Received: 26 July 2024 / Accepted: 16 April 2026 / Published online: 28 May 2026

Abstract

Storyboards are one of the most effective methods to pre-visualize a story: a sequence of illustrations or pictures portraying settings, characters, actions, and events. This graphical representation is a roadmap that guides industry professionals through the story. The storyboarding process is a key step for both movie and video game productions, allowing creators to visualize the scenes and find potential problems before they occur. While storyboards are usually drawn or sketched in 2D, 3D approaches enable storytellers to achieve a better understanding of spaces, objects, lights, and camera movements: these approaches are usually named "3D Storyboards". The feeling of immersion provided by extended reality (XR) technologies can further enhance these benefits. This manuscript investigates and compares two 3D XR-based storyboarding solutions, which share a common graphical user interface. The proposed systems have been tested with domain experts and students in a tabletop-scale scenario. Experimental results clearly indicate that the virtual reality-based solution, with the current interaction standards imposed by the chosen systems, outperforms the augmented reality in terms of usability; moreover, participants reported a lower workload when using an input interface based on physical controllers.

Keywords Storyboarding · Virtual reality · Augmented reality · User study

1 Introduction

A storyboard is a sequence of images visually representing a story by depicting settings, characters, actions, and events. Storyboards are one of the most effective methods to pre-visualize a story, helping creators to visualize the scenes and find potential problems before they occur during production. The storyboarding process is a key step in both movie and video game productions, but it is also widely applied in other domains, including theater performances, advertising campaigns, virtual project presentations, and software development, particularly for describing user interfaces and user interactions.

A sample of a storyboard sequence is presented in Fig. 1: in this example from the video game Ryzom, for each shot, the image is depicted on the left, actions (and possibly dialogues) are detailed under the image, and shot numbers and other technical details are provided on the right. Storyboards can provide diverse levels of detail depending on different templates. The key elements of a storyboard are:

- Shot images: panels able to show characters and their actions
- Shot numbers: numbers indicate the temporal position of each shot in the current scene
- Action(s): a description of what happens in a shot can also be included
- Dialogue(s): conversations between two or more characters and/or background narrations

✉ Federico Manuri
federico.manuri@polito.it

Francesco De Pace
francesco.de_pace@cim40.com

Valentina Carrescia
valentina.carrescia@gmail.com

Sara Piumatti
sara.piumatti@gmail.com

Andrea Sanna
andrea.sanna@polito.it

[1] Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

[2] Digital Factory, Competence Industry Manufacturing 4.0, Corso Settembrini 178, 10135 Torino, Italy

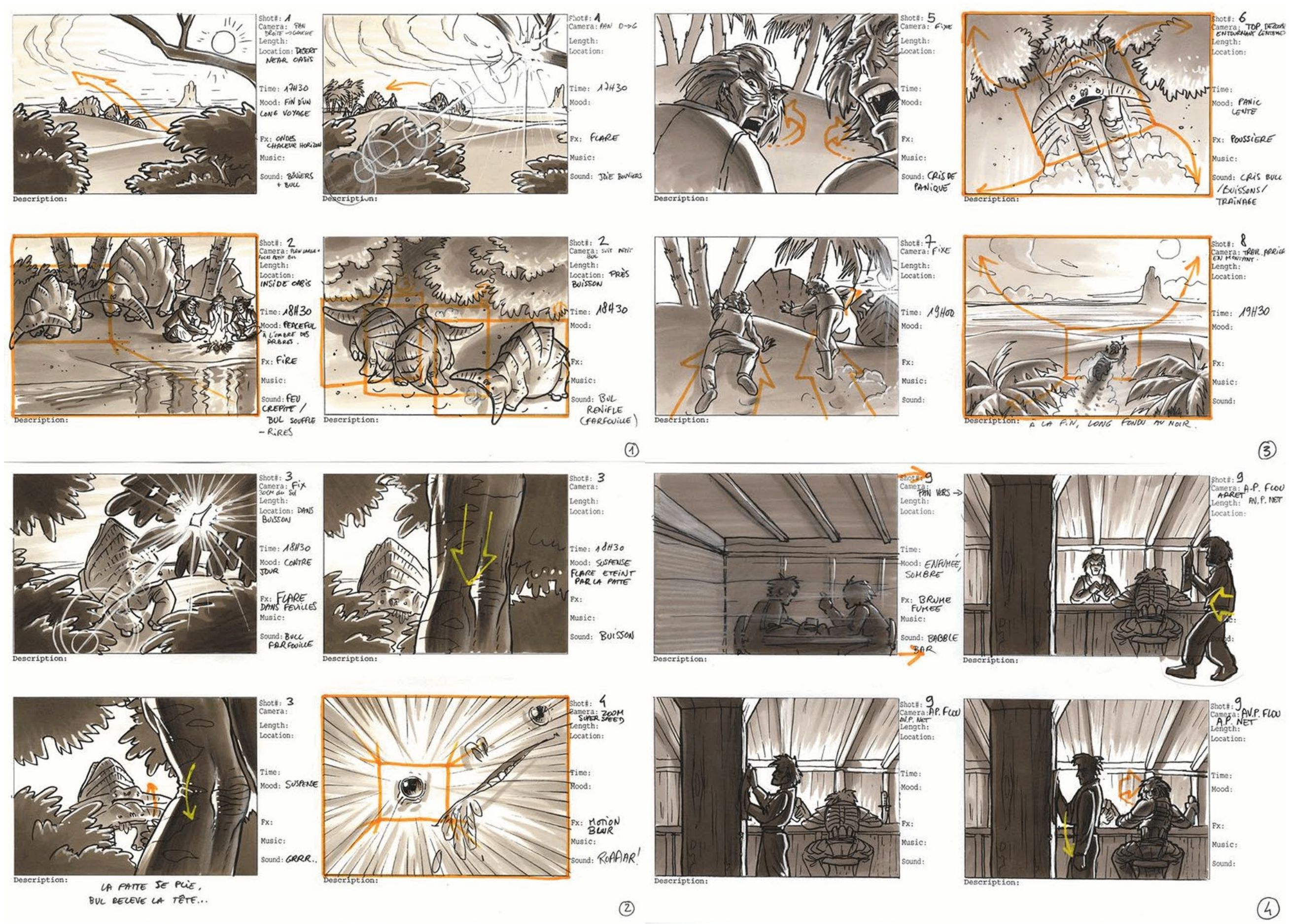

Fig. 1 A sample of a storyboard sequence from the open-source video game Ryzom; image by Ryzom available under CC-BY-SA−2.0 License. For each shot, the image is depicted on the left, actions (and possibly dialogues) are detailed under the image, and shot numbers and other technical details are provided on the right

- Technical details: these details explain how the shot should be created in production; for instance, camera parameters and sound effects are often included, as well as camera movements

Storytellers (also referred to as storyboard artists in the rest of the paper) usually draw or sketch storyboards in 2D, either with traditional or digital tools (Adobe 2025). While this process may appear intuitive, it requires good drawing skills and the ability to visualize how the camera will frame the scene before it is drawn. Picture-based applications help overcome the need for drawing skills by offering datasets of pre-made images that users can incorporate into their storyboards (Clever Prototypes 2025; Canva 2025). However, users have to learn how to use the application, search for appropriate images for each panel, and work within the constraints of the available dataset. In 3D animation productions, artists have to convert from 2D storyboards to 3D scenes and assets. However, since it is very difficult to approximate a real camera shot when drawing a 2D panel, this transition might lead to inconsistencies, such as objects occluding the camera or incorrect scale factors for set elements (Gouvatsos 2018).

In recent years, 3D game engines, modeling, and animation tools have been researched and developed, and novel solutions have emerged to support the storyboarding process in 3D (Galvane et al. 2019; Rao et al. 2023), addressing the limitations of the 2D approaches. The resulting storyboard panels can depict the relative positions of objects and characters within the virtual scene, along with their spatial relationships (Siddi 2025). A key advantage of a 3D approach is long-term cost reduction, since tasks such as adjusting camera positions, modifying lights, or adding simulations can be automated. Therefore, 3D storyboarding applications address the challenge of converting 2D shots into 3D layouts, offering artists and filmmakers an interactive pre-visualization (Gouvatsos et al. 2014). Using 3D tools (e.g., game engines), storytellers can work directly with 3D resources to precisely define what will appear on the screen (Chapman 2023). Moreover, 3D solutions support storyboard artists by providing immediate feedback on how animations and actions will appear (Lin et al. 2015; Trottnow et al. 2015); in particular, game engines can deliver real-time responses with high production quality (Moran 2015; Kilkenny 2023). As a result, the length of a panel can be specified by computing the time needed to play the corresponding actions (see Fig. 2 as an example), thus providing an intermediate product between a simple "static" storyboard and an animatic. This approach allows potential issues to be detected and fixed in the early stages of the production pipeline.

Recent advances in extended reality (XR) demonstrate how 3D desktop applications may benefit from such technologies. Augmented reality (AR) and virtual reality (VR) offer storytellers new possibilities that go far beyond traditional 3D environments (Nebeling et al. 2021). XR technologies can transform and enhance the perception of the space, extend the user experience, and foster social interaction by enabling users to share digital spaces. It has been shown that VR tools enhance a storyboard by giving more control of the camera and animated assets (Schaefer 2024), and they may be used to create flattened images for traditional storyboards. Although XR technologies for storyboarding have been explored in the past, a comparison between different 3D XR-based storyboarding systems has not yet been conducted. Existing AR and VR commercial solutions present many differences and can benefit end users in distinct ways. It is therefore important to determine whether, given the same UI, scenario, and task, users show a preference for one of the two solutions. This paper focuses on the 3D storyboard domain (hereinafter referred to simply as "storyboard"). In particular, the design, development, and comparison of two 3D XR-based applications is performed. Comparing an AR-based and a VR-based storyboarding system should provide valuable insights to foster XR storyboarding research and highlight possible future trends. The aim of this research is to answer the following research questions:

- **R1a**: Do users prefer the interactions of an augmented or virtual reality commercial system for creating 3D storyboards in a tabletop scenario? **R1b**: Why do they prefer one or the other application?
- **R2**: What is the usability perceived by the users with the two solutions?
- **R3**: What is the level of workload perceived by the users with the two solutions?

In this research, AR and VR storyboarding apps have been designed and developed for two wearable devices; in particular, the Microsoft HoloLens 2 has been selected as the AR device, and the Oculus Rift as the VR headset. The tabletop scale has been chosen for both applications to migitate the issues due to the narrow field of view (FoV) of AR optical see-through devices. The graphics user interfaces (GUIs) have been designed by leveraging the Mixed Reality Toolkit (MRTK), thus minimizing the differences between the AR and VR applications.

A three-step approach has been chosen to compare the two storyboarding applications. First, to unearth and fix usability problems, we tested the applications with usability experts. We then gathered domain expert feedback from two professional storyboard artists in a simulated storyboarding session. Finally, we performed a between-subjects statistical evaluation of the proposed solutions with students; two separate user groups evaluated the applications to avoid learnability effects due to the similarities between the two solutions. Users created a storyboard, either in AR or VR, and then evaluated the system by filling out two questionnaires to measure usability and task workload. Furthermore,

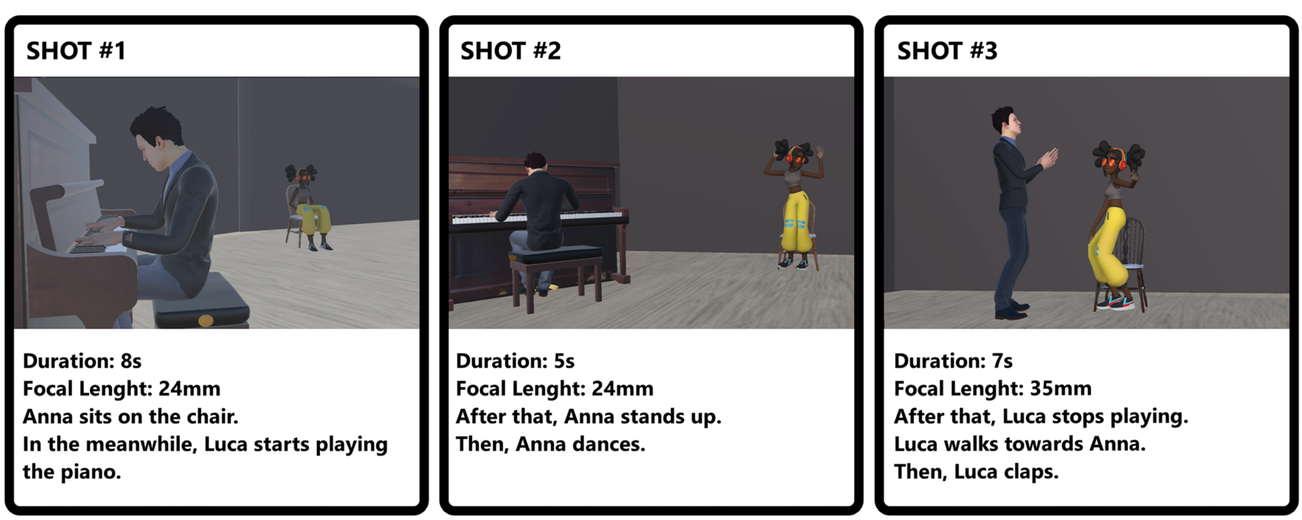

Fig. 2 An example of a storyboard that shows the story of two characters: Anna and Luca. Shot numbers, scene lengths, focal lengths, and text descriptions are associated with each image of the sequence

the users answered four open questions to provide further insight with regard to the proposed system, perceived limitations, and possible improvements. Experimental results clearly indicate that the virtual reality-based solution, with the current interaction standards imposed by the chosen systems, outperforms the augmented reality in terms of usability. Moreover, participants reported a lower workload when using an input interface based on physical controllers.

The paper is organized as follows: Sect. 2 reviews different approaches for 3D storyboarding, and XR benefits are also discussed. The design and implementation of the two XR solutions are described in Sect. 3. User tests and obtained results are presented in Sect. 4 and discussed in Sect. 5.

2 Related works

Different solutions to create digital storyboards based on 3D scenes have been explored in previous works, supporting the growing interest in novel storyboarding approaches. Thus, a thoughtful analysis is necessary to obtain a clear insight into the pros and cons of existing solutions. The following sections present both an analysis of 3D storyboarding research, exploring different approaches to creating storyboards, and XR-based solutions, which investigate the benefits of AR and VR.

2.1 Traditional 3D storyboarding approaches

Storyboard panels should depict characters and their actions. In a 3D environment, the user has to set up the stage and define the characters' poses. The stage is either obtained by modeling the assets or loading them from a library; the first approach corresponds to drawing the scene in a 2D storyboard, whereas the second approach is similar to picture-based applications. Character posing is a fundamental step in the process of 3D animation (Sito 2013), as characters' poses usually represent key moments of a story. As for the stage setup, it is possible to manually define poses with a forward and inverse kinematics approach or rely on animation libraries with a scene enactment approach.

2.1.1 Forward and inverse kinematics

2D animations simulate motion through the rapid succession of 2D images with slight differences; 3D animations are based on changing the position and rotation (or pose) of the 3D models in the virtual space in a given time frame. For complex objects such as characters, a skeleton is usually defined and rigged to the 3D mesh; the user can specify the character's pose by moving and rotating the skeleton bones both in forward and in inverse kinematics. A typical human-character model is represented by more than 50 degrees of freedom, whereas a typical user input is way more low-dimensional (Wei and Chai 2011). Thus, an easy-to-use interface for 3D character posing should support various user constraints to empower users with different skill levels: professional users should be able to curate character poses in detail, whereas less experienced users should be able to define realistic poses without too much effort. Unfortunately, traditional 3D editing tools remain complex and laborious compared to 2D sketching; virtual character posing is a time-consuming and skill-intensive task that requires animators to work with sophisticated user interfaces. Commercial solutions, such as Frameforge 3D (Innovative Software 2024) or Shot-Pro (WebGames3D com 2024), usually target expert users, as they are complex to master and present a steep learning curve.

Due to the complexity of traditional character posing, other research works investigate the usage of custom tangible interfaces instead of traditional input modalities to simplify the posing process. Ishii and Ullmer first researched tangible interfaces to replace the mainstream skeleton manipulation process (Ishii and Ullmer 1997) since mouse and keyboard interfaces fall short because their control spaces do not match the perceptual space of the 3D interaction task (Jacob and Sibert 1992). Jacobson et al. (2014) proposed a tangible device for direct, physical manipulation with degrees of freedom in direct correspondence with the 3D rotations at the skeletal joints of the virtual character. The user may quickly assemble a custom device based on modular, hot-pluggable mechanical parts to establish a one-to-one mapping between the physical device and a virtual skeleton. However, professional character rigs may contain hundreds of bones, spanning many levels in the skeletal hierarchy. Thus, with this approach, the user would need a lot of parts and time to model a complex rig or limit the interface mapping with partial binding at a specific level of the rig's hierarchy. This limitation was addressed by Glauser et al. (2016), introducing a symbiotic software and hardware approach to directly control complex rigs using only a small set of physical controls to avoid both oversimplification of the pose space and excessively bulky device configurations. However, the system's capability to create realistic poses highly depends on the retargeting algorithms and the pose interpolation schemes available to the user.

2.1.2 Scene enacting

A less explored approach toward 3D character posing consists of enabling the user to enact the actions depicted in the script. Pizzi et al. (2010) proposed an authoring tool allowing game designers to formalize, visualize, modify, and validate game-level solutions in the form of automatically generated storyboards. The proposed system automatically generates possible solutions to a given game level using the player character as the main agent and gameplay actions as the basic elements of solution generation, using the description of the initial state, the level's objective, and the level layout as input. The storyboard generation algorithm runs parallel to the solution generation phase, using templates to construct each storyboard panel from data corresponding to the actions performed to solve the game level. The panels are then assembled sequentially into a full storyboard presenting the complete level solution generated by the system. However, this solution is limited in terms of user customization, similar to automatic 2D to 3D or text to 3D solutions.

More recently, Manuri et al. (2024) explored a similar approach to automatically generate a storyboard that gives users complete control over the actions performed by the characters. This system enables users to select the available characters or the camera, control them in first person, position them in the scene, and perform actions chosen from a list (each action corresponds to a character animation). The system can automatically generate the list of available actions from the context (e.g., characters can open a door only if they are close enough), relying on the concept of state-machine. Finally, the descriptions for each storyboard panel are automatically generated: the sequence of the performed activities is used to create the text descriptions, whereas screenshots are taken from the point of view of the virtual cameras. This approach overcomes the need for specific user skills, thus providing a simple point-and-click interface. Moreover, it confers complete control to the user, minimizing the constraints on the artist's creative process, which is free to explore visual alternatives starting from the original script. The limitations of this approach in a traditional 3D desktop application involve difficulties in camera usage and character and object selection, which other 3D assets may obscure. However, an immersive VR solution would easily overcome these limitations, as further discussed in Manuri et al. (2023).

2.2 Alternative 3D storyboarding approaches

Based on the state-of-the-art analysis, three research approaches have been explored to improve 3D storyboarding methods: 2D to 3D conversion, text to 3D conversion, and storyboarding in XR.

2.2.1 2D to 3D sketch conversion

2D sketching is part of most animation pipelines and is used by both animators and pose artists to define the initial pose before the animation stage. One approach to pose 3D characters consists of drawing 2D sketches and automatically translating them into 3D poses, with the simplest solution being the analysis of stick figures (Xiao et al. 2015). However, 2D sketches do not provide any volume data; thus, depth and symmetry ambiguities may arise. To address this problem, systems may require additional input data in the form of silhouettes (Hahn et al. 2015), lines of action (Guay et al. 2013), or gesture drawings (Bessmeltsev et al. 2016). However, additional input may be more complex to analyze in real-time, require more effort by the user creating the sketch, or require multiple refinements by the user. The solution proposed by Mao et al. (2005) requires the user to draw lines with different thicknesses to distinguish between body parts closer to the camera and those far away. Interactive systems may require the user to refine progressively the character pose through body lines (Mahmudi et al. 2016); Mouajjah and Plessiet proposed a progressive incremental method that lets the user specify the overall masses in the scene, define the initial pose by body lines and sticky figure lines, and build the character pose gradually by adding finer details (Mouajjah and Plessiet 2021).

Automatic solutions have also been explored to infer a 3D pose from a drawing, matching the 3D model projection and inferring the missing depth. This result can be obtained with a database approach, using computer vision techniques to compute the tracking and posing (Jain et al. 2009) and positioning the character in the scene (Xu et al. 2013). An alternative to the database approach is the explicit shape model approach, which compares the 2D image with a rendered pose of a 3D model to estimate how close it is to the correct pose (Favreau et al. 2004). This approach is efficient if multiple camera views are available, but requires a 3D model similar to the one represented in the 2D image. Gouvatsos et al. (2017) formulated the inference of the 3D pose as an optimization problem, proposing a parallel variation of the Particle Swarm Optimization algorithm called PARAC-LOAPSO for searching the minimum.

Overall, approaches based on the 2D to 3D sketch conversion require the user to have good drawing skills. Moreover, automatic approaches are less flexible with regard to the artist's creative process, whereas interactive approaches require the user to learn how to convey depth information through additional inputs or sketching techniques.

2.2.2 Text to 3D conversion

Another approach investigated to improve the storyboarding pipeline focuses on using text and annotations to convey additional information from the script automatically. Louarn et al. (2018) propose a staging specification language that receives as inputs a 3D model of the scene with the environment, props, and characters, and a staging specification describing the position of cameras and characters in the 3D scene. The system outputs a scene layout with characters and cameras placed following the staging specification. Zhang et al. (2021) propose a 3D animation system that allows novice users to create, edit, preview, and render animations. The pipeline is based on text editing, parsing the texts as semantic scene graphs through natural language processing, retrieving 3D object models for virtual scene composition, and applying motion clips for character animation. Kim et al. proposed ASAP (Kim et al. 2021), a tool for auto-generating storyboards and previz for screenwriters and filmmakers. The system expects a script formatted following the Final Draft standard, then parses it into actions, characters, and dialogues by combining deep learning, data-driven, and rule-based approaches. Based on these inputs, it generates virtual humans' physical motions and co-speech gestures, presenting natural behavior/dialogue scenes. Recently, Rao et al. (2023) proposed the Virtual Dynamic Storyboard (VDS), allowing users to storyboard shots in virtual environments. The system requires a formatted story script and a camera script as input to generate several character animations and camera movement proposals following predefined story and cinematic rules to allow an off-the-shelf simulation engine to render videos.

Whereas these approaches are very promising for automating part of the storyboarding process, existing solutions present two limitations: they require a custom textual input, either providing additional annotations or specific formatting to the original script; they do not allow the user to customize the final result by an approach that is not text-based, thus implying less flexibility with regard to the artist's creative process.

2.2.3 3D storyboarding in extended reality

Previous research studies explored the usage of XR to benefit the storyboard and animation process. One of the first examples of AR storyboarding tools was investigated by Shin et al. (2005): the proposed system allowed the user to place physical markers on a desk to define the scene composition and the characters' poses. The project aimed to exploit AR to enable non-experienced users to compose 3D scenes for creating storyboards in a tabletop real environment. Using ARToolKit (Kato and Billinghurst 1999) and an RGB camera tracking the desk, the system could detect and convert the markers into 3D assets and compose the 3D scenes for a storyboard.

With the diffusion of Immersive Virtual Reality solutions, traditional storyboarding tools have proven to be shallow and restrictive with regard to the challenges posed by planning and producing fully explorable 3D environments. Henrikson et al. investigated the problem of creating movies for VR, collaborating with film professionals to explore and address creating storyboards for VR and the related challenges (Henrikson et al. 2016): the proposed system pairs a sketch tablet and a VR head-mounted display (HMD), allowing one user to sketch a storyboard scene on stacked panoramic panels, corresponding to concentric cylindrical surfaces of increasing view depth, whereas the other user can visualize the resulting scene in VR. The proposed prototype is a collaborative, multi-device storyboarding solution that supports quick authoring and fast virtual environment navigation.

Berford et al. (2017) addressed the limitations of animating in traditional third-party tools and importing into VR, and the issues related to the context switch between VR and 2D monitors. They proposed a fully VR animation layout tool, Dollhouse, that allows users to position 3D assets in a scene, set keyframes, and playback the state of the animation. They further expanded this tool with Maestro, a cross-platform sequence management and playback tool that allows any number of users to review in VR a 3D experience. The sense of depth allows the viewer to mentally navigate through the space, clearly understanding how scene elements are laid out and how the user may navigate the scene.

Galvane et al. (2019) explored the usage of VR as a content creation tool for movie pre-visualization, exploiting the ease of use of VR manipulations for scene composition and character animation. The proposed system also offers a storyboarding mode, which allows users to define the static scene, the character poses (with a character posing approach, as described in Sect. 2.1.1), and the cameras, annotate each shot, and specify camera motions. The proposed tool aims to rely on VR immersive technologies to author animated 3D content, exploiting the capacity to manipulate virtual objects using 6D input devices.

Manuri et al. (2025) recently proposed a novel paradigm to create storyboards in XR in a 1:1 scale environment. The proposed approach leverages the capabilities of object-detection and pose-estimation systems to include the physical environment and all its furniture in the storyboard, exploring the interaction between virtual characters and physical objects to define the character poses. This research represents a starting point for enacting scripts and framing the environment to obtain storyboard panels that mix real and virtual elements.

From this analysis, an XR storyboarding application that relies on libraries of 3D assets and animations would overcome the limitations of other methods:

- It would not rely on the user's drawing or posing skills
- It would not require custom inputs nor the user's skills to create them

- It would offer higher flexibility with regard to the artist's creative process and customization capabilities

3 Materials and methods

From the analysis of the state-of-the-art, it is possible to summarize the features provided by the different storyboarding approaches and paradigms, as depicted in Table 1. The following sections present the requirements, the proposed application workflow, and the development of the two XR solutions.

3.1 Requirements

As depicted in Table 1, 3D approaches improve on 2D methodologies as the storytellers can achieve a better understanding of spaces, objects, lights, and camera previews. To design and develop the two XR applications, it is necessary to define the functional requirements. Truong et al. (2006) defined five essential attributes that characterize a storyboard as:

- **Req1**: level of detail in terms of elements depicted in the frame and photo-realism of the representation
- **Req2**: inclusion of text describing the story for each panel
- **Req3**: inclusion of people and emotions to show interactions, motivations, or to build empathy
- **Req4**: number of frames used to depict the story
- **Req5**: indicating the time passing or conveying changes over time

Additionally, to further enhance the benefits provided by 3D solutions, XR technologies should provide:

- **Req6**: feeling of immersion (Henrikson et al. 2016; Berford et al. 2017; Galvane et al. 2019), allowing the user to experience the scene from the characters' point of view and explore different storytelling options.
- **Req7**: easy camera movements (Berford et al. 2017; Galvane et al. 2019; Manuri et al. 2024, 2023)
- **Req8**: easy environment navigation (Berford et al. 2017; Galvane et al. 2019)
- **Req9**: natural object manipulation (Galvane et al. 2019; Manuri et al. 2024, 2023)
- **Req10**: multi-user support (Henrikson et al. 2016; Berford et al. 2017)

Finally, to overcome the limitations depicted in the related work analysis, the application should

- **Req11**: not require the user to have good drawing skills (Xiao et al. 2015; Hahn et al. 2015; Guay et al. 2013; Bessmeltsev et al. 2016; Mao et al. 2005; Mahmudi et al. 2016; Mouajjah and Plessiet 2021; Jain et al. 2009; Xu et al. 2013; Favreau et al. 2004; Gouvatsos et al. 2017; Henrikson et al. 2016)
- **Req12**: not require the user to have experience in 3D animation and posing techniques (Xiao et al. 2015; Hahn et al. 2015; Guay et al. 2013; Bessmeltsev et al. 2016; Mao et al. 2005; Mahmudi et al. 2016; Mouajjah and Plessiet 2021; Jain et al. 2009; Xu et al. 2013; Favreau et al. 2004; Gouvatsos et al. 2017; Henrikson et al. 2016; Louarn et al. 2018; Zhang et al. 2021; Kim et al. 2021; Pizzi et al. 2010; Manuri et al. 2024, 2023)
- **Req13**: not rely on a preformatted script or additional inputs to the system (Louarn et al. 2018; Zhang et al. 2021; Kim et al. 2021)

Based on these requirements and the features detailed in Table 1, the scene-enacting approach has been chosen for the development phase of the two XR applications.

3.2 The application workflow

The proposed storyboard workflow is depicted in Fig. 3. Firstly, the user defines the reference table according to the chosen technology. Then, an XR storyboarding solution requires two phases:

1. **Virtual Stage Setup:** the user builds the virtual set for a scene, selecting 3D assets from a virtual library and placing them in the virtual environment
2. **Storyboarding:** the user poses the characters with regard to the script, sets the lights, and positions the camera to frame the scene; after framing the scene, the system will automatically generate the text describing the panel and add the current panel to the storyboard

3.2.1 Virtual stage setup

The first step consists of building the virtual set for a scene. A drag-and-drop approach can foster the creation of a 3D set, allowing the user to select 3D assets from a virtual library (**Req1**, **Req11**). When the set is ready, the user can define the starting position for the characters, using the same approach proposed for setting up the 3D stage.

3.2.2 Storyboarding

The storyboarding phase requires the user to set up and frame the scene, defining several shots to determine the

Table 1 The storyboarding features provided by the different 2D and 3D approaches discussed in the literature

paradigm	approach	real camera preview	proper space perception	realistic spatial relationships among objects	support multiple cameras	drawing skills not required	easy lights and shadows management	simulations	character pose flexibility	posing skills not required
2D	drawing (Adobe 2025)	X	X	X	X	X	X	X	✓	✓
	picture-based (Clever Prototypes 2025; Canva 2025)	X	X	X	X	✓	X	X	X	✓
3D	posing (mouse) (Innovative Software 2024; WebGames3D com 2024)	✓	✓	✓	✓	✓	✓	✓	✓	X
	posing (tangible) (Ishii and Ullmer 1997; Jacob and Sibert 1992; Jacobson et al. 2014; Glauser et al. 2016; Shin et al. 2005; Kato and Billinghurst 1999; Berford et al. 2017; Galvane et al. 2019)	✓	✓	✓	✓	✓	✓	✓	X	✓
	2D to 3D (Xiao et al. 2015; Hahn et al. 2015; Guay et al. 2013; Bessmeltsev et al. 2016; Mao et al. 2005; Mahmudi et al. 2016; Mouajjah and Plessiet 2021; Jain et al. 2009; Xu et al. 2013; Favreau et al. 2004; Gouvatsos et al. 2017; Henrikson et al. 2016)	✓	✓	✓	✓	X	X	X	X	✓
	text to 3D (Louarn et al. 2018; Zhang et al. 2021; Kim et al. 2021)	✓	✓	✓	✓	✓	X	X	X	✓
	scene enacting (Pizzi et al. 2010; Manuri et al. 2024, 2023)	✓	✓	✓	✓	✓	✓	✓	X	✓

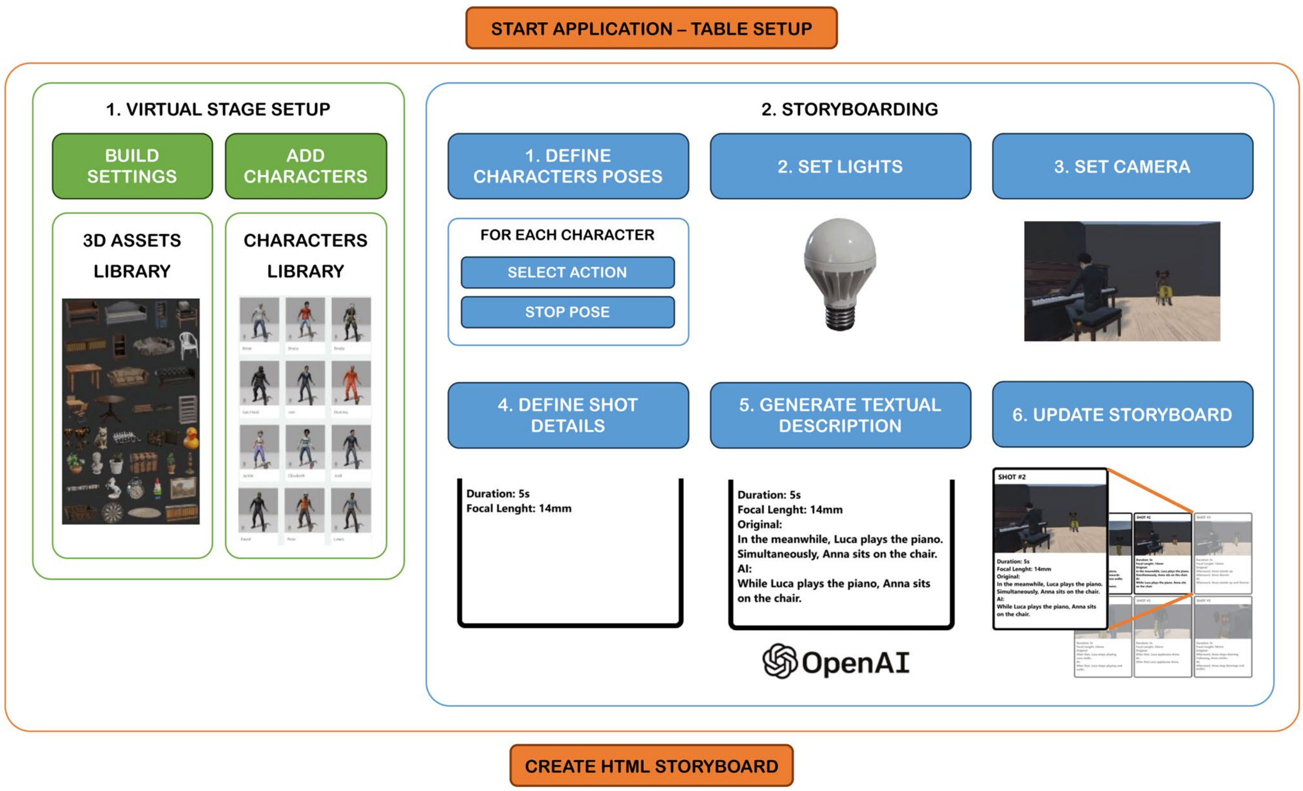

Fig. 3 The proposed workflow to generate 3D storyboards in extended reality

storyboard based on the script. To this end, the user has to repeat the following actions for each shot:

1. **Define Characters Poses**: the user selects the characters, one by one; for each character, one or more actions can be chosen; upon selection, the corresponding animation(s) starts; the user stops the animation(s) to define the character pose (**Req12** and **Req3**). Actions can target the character itself, another character, or an object in the scene. Actions not dependent on another 3D element are displayed when a character is selected; selecting other 3D elements, actions that the current character can perform on those targets are shown. It is relevant to clarify that actions can be cyclical or one-shot. Cyclical actions are started by selecting the corresponding action, and they only stop when a different action is selected. One-shot actions start and stop autonomously, and then the character returns to the previous pose. For instance, a standing character can jump and, at the end of the animation, will be standing again. Moreover, the user can specify if the actions pertaining to one shot are concurrent or consecutive; therefore, different temporal adverbs will be used when generating the phrase based on the user's choice (**Req5**). Finally, some actions may allow the user to specify additional details with a virtual keyboard. For example, if the user selects the 'talk to' action, the system shows a prompt to type the dialogue using the virtual keyboard or to dictate it. The resulting text will be automatically included in the panel's textual description. If the action described in the script is not available in the action database, the user can select the 'other' action option: a prompt to digit the action name will appear to generate the shot description properly, whereas the character pose corresponding to the action will not be available, and the character pose will not change
2. **Set Lights**: when the user completes the posing step for each character, lights in the scene have to be set: position, direction, intensity, and color have to be specified (**Req1** and **Req11**)
3. **Set Camera**: the user can then place one or multiple cameras in the scene and try out different camera's parameters, trying out different shots to depict the current storyboard panel (**Req1** and **Req11**)
4. **Define Shot Details**: the user can specify additional details that will be part of the shot description in the storyboard, e.g., the shot duration (**Req4**)
5. **Generate Textual Description**: after that, the system automatically converts the actions performed by the characters into a textual description based on the character name, the action name, and (eventually) the target name (**Req2** and **Req13**). The resulting text can be improved by performing a rephrasing step based on

generative pre-trained transformer models fine-tuned to target conversational usage, such as ChatGPT.

6. **Update Storyboard**: the resulting storyboard panel, including the shot, the text, and the camera details (such as the focal length), is added to the storyboard.

This process is repeated for each shot of the current scene, whereas the user has to build a new virtual setting when moving on to another scene of the script. When the last panel of the storyboard is ready, the system can generate the final storyboard as an HTML file that can be further reviewed with the opportunity to edit the textual content or reorder the panels.

This workflow's scalability solely depends on the 3D asset and character libraries. Existing productions already have tons of 3D content that can be used to populate such libraries, whereas websites providing free 3D models and animated characters exist, even with animations (e.g., Mixamo, Adobe 2024).

3.3 The XR applications development

To effectively and fairly compare two XR systems that rely on different visualization and interaction approaches, it is necessary to develop two applications that could be experienced similarly with both XR paradigms. Thus, the proposed functionalities and the UI elements should be almost the same. On the other hand, the visualization imposed by the chosen systems (AR or VR) and interaction paradigms (hand or controller), which are representative of the chosen XR systems, should be the only differences between the two systems.

3.3.1 Technologies

The Unity game engine (Unity Technologies 2024) has been selected to develop the proposed system since it allows developers to deploy the resulting application for most of the AR and VR devices available on the market (**Req6-Req10**). The HoloLens 2 by Microsoft has been chosen as the target AR device since it represents the state of the art in terms of wearable see-through devices. The VR device of choice is the Oculus Rift headset: it is a valuable device for tabletop VR experiences and is easier to set up than the HTC VIVE headset. Moreover, the Mixed Reality Toolkit 2 (MRTK2) (Microsoft 2024) has been chosen to design the GUI: this Microsoft-driven project provides a set of components and features tailored for cross-platform Mixed Reality app development in Unity, thus providing the cross-platform input system and building blocks for spatial interactions and UIs. To complete the system, the proposed software architecture relies on OpenXR (Group 2024), an open standard that provides cross-platform, high-performance access to XR platforms and devices, enabling applications and engines to run on any system that exposes the OpenXR APIs.

3.3.2 The user interface

The GUI has been designed to provide an easy, homogeneous interface with a low learning curve; this GUI could be effectively adopted by users not proficient in the AR or VR domains. The main menu displays the core options available to the user based on the application status; options are:

- Creating a new scene
- Loading or saving an existing scene
- Renaming a scene
- Running the tutorial
- Switching between the virtual stage setup and the storyboarding
- Adding a new camera
- Taking a screenshot
- Finalizing the storyboard

During the 3D scene setup, a 3D asset library is shown to the user. The user can select objects to build the setting and add characters. A piano, a chair, three walls, and a floor are the setting elements, whereas a female (Anna) and a male (Luca) are the two available characters, as depicted in Fig. 4. The selected (or active) character is Luca, as outlined by the red circle around Luca's feet.

During the storyboarding, the user sees: the list of actions available for the active character, a menu with the option affecting the current vignette of the storyboard, the automatically generated text description, and the screenshots rendered from the point of view of the active virtual camera, which represent the storyboard vignettes. A 3D camera is added by default to the scene to allow the user to take the screenshot; the main menu allows the user to add multiple cameras to the scene. The camera has a detachable panel representing the camera screen, thus enabling the user to see what the camera frames independently of the user's position; this panel can be placed wherever in the 3D space.

3.3.3 User input

The system has been designed to foster a natural interaction with the GUI and the 3D objects. A point-and-click approach is provided to interact with faraway virtual elements. With the Oculus Rift, a virtual ray is cast by each hand controller; users can move the controllers to point the rays in the virtual environment as they would with the mouse in a 2D application. The click is performed with the

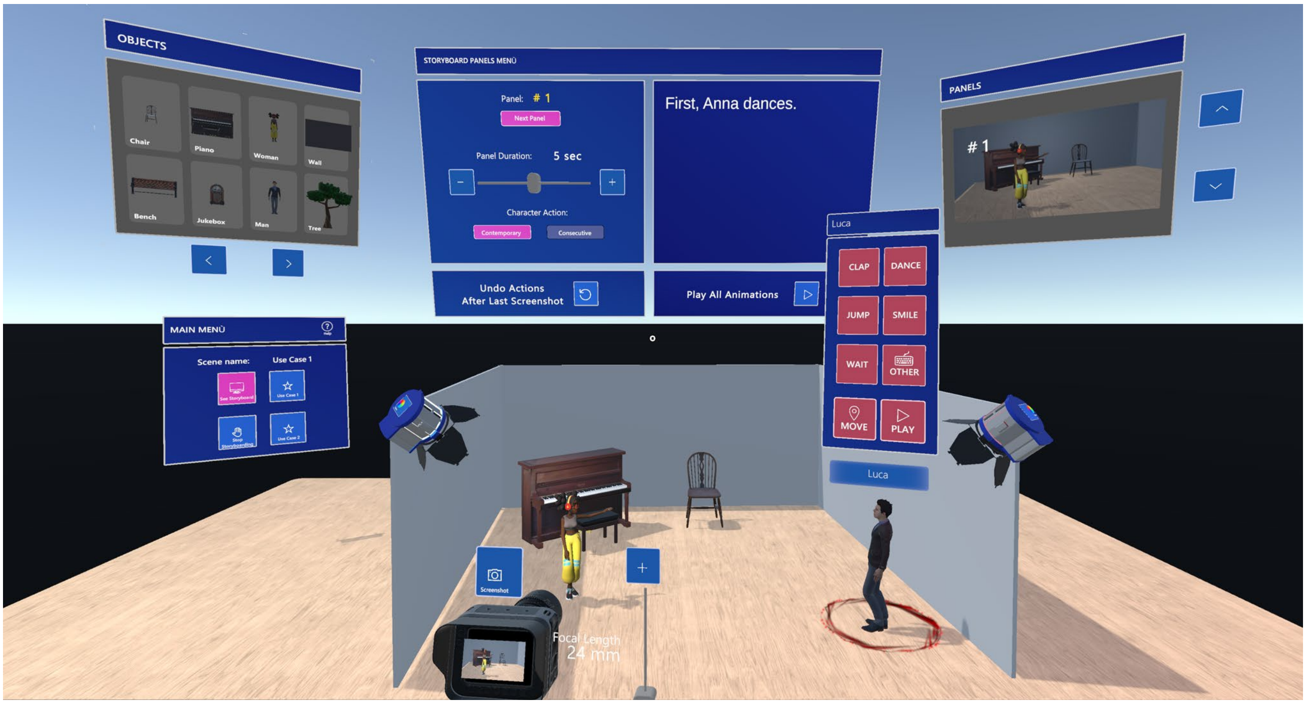

Fig. 4 The UI. All the elements of the interface are displayed: settings, characters, panels, virtual cameras, the taken screenshot, the text description, and so on

trigger button. With the HoloLens 2, users have been given two options: a hand-tracking input profile or a gaze-tracking and voice recognition profile. The first option is equivalent to the Oculus Rift approach, but the hand is tracked instead of the controller, and the click is performed with the pinch gesture, thus relying on the MRTK2 Gesture Recognition module. The second approach depends on the gaze-tracking capabilities of the HoloLens 2. It allows users to control the cursor position by moving their head, whereas the click is performed by pronouncing the word ‘select’, thus relying on the MRTK2 Windows Speech Input module. Moreover, a near-interaction modality was provided for both applications. With the Oculus Rift, users can interact with buttons and other ‘pushable’ elements simply by touching them with the controllers, whereas ‘graspable’ objects can be moved around by merely pressing the grip button. With the HoloLens 2, the same near-interaction modality as the Oculus Rift can be performed by touching virtual objects with the index finger or using the pinch gesture on nearby objects to grab them.

4 Evaluation

A three-step approach has been chosen to compare the two storyboarding applications:

1. The first step involves experts in testing usability
2. Then, professional storyboarders tried the application and gave feedback
3. Finally, students tested the application, and their feedback was collected by questionnaires

4.1 Expert usability evaluation

Three usability experts evaluated the AR and VR applications separately. The scope of this preliminary evaluation of the two XR applications is to unearth the most relevant problems and to fix them before performing the other two evaluation steps.

4.1.1 Metrics

Feedback from experts was collected in two ways:

- Comments provided while testing the application, to highlight usability issues that should be fixed
- A written evaluation based on Nielsen’s heuristics (Nielsen 1994); experts were asked to rate the system for each heuristic on a 1-to-5 Likert scale, with one corresponding to “Severe usability issues" and five corresponding to “Free from usability issues"; experts were also asked to describe the usability issues they identified for each heuristic

Nielsen and Molich (Nielsen and Molich 1990) showed that aggregating evaluations from three to five evaluators is an effective approach to address most of the usability problems of a system/interface.

4.1.2 Participants

Usability experts have been selected among scholars from the Department of Control and Computer Engineering of the Politecnico di Torino, whose research is mainly related to human factors and extended reality, and who have at least three years of experience in this research field.

4.1.3 The selected use case

The use case consists of creating a storyboard using the following short script as a reference for testing all the application's functionalities:

```
  Anna and Luca are in their home.
  Luca approaches the piano and begins
to play, while Anna sits next to
  him.
  At a certain point, Anna gets up from
the chair and starts dancing.
  Luca then stops playing and approaches
Anna.
  Luca applauds Anna.
  Anna stops dancing and smiles at him.
```

Users can decide how many panels to use to depict the actions described. The script was available for inspection anytime in the AR/VR UIs.

4.1.4 The test procedure

The overall test procedure can be summarized as follows:

- Introduction: the user is introduced to the proposed research, and the test procedure is explained in detail
- User data: the user fills out a preliminary form used to collect general information about participants
- Test: the user is introduced to the application by a tutorial, then creates a storyboard based on the proposed use case. In this phase, the user is asked to provide feedback about usability issues as soon as they arise
- Questionnaires: the user has to score the ten Nielsen's heuristics on a 5-point Likert scale, based on the problems identified while testing the application

4.1.5 Results

Table 2 details Nielsen's heuristics scores provided by the experts. The two applications provide similar usability, with a delta (Δ) value between the average scores lower than one for seven out of ten heuristics. Moreover, all the heuristics were rated with an average score of at least three, confirming that only minor usability issues arose during the tests. All the issues about heuristics with an average score of less than four have been addressed to improve usability. A complete list of the issues and suggestions detailed by the experts is provided as supplementary materials.

4.2 Professional storyboarder evaluation

Professional storyboarders participated in the user study to assess both strengths and weaknesses of the proposed workflow. This evaluation step aims to determine whether the proposed system can effectively support storyboard artists and other experts involved in the filmmaking and animation pipelines.

4.2.1 Metrics

After testing the two applications, users were asked to provide their feedback in an interview, highlighting their pros and cons.

Table 2 Nielsen's heuristics scores provided by the experts (E1-E3)

		AR				VR				
#	Nielsen's Heuristic	E1	E2	E3	AVG	E1	E2	E3	AVG	Δ
1	Visibility of System Status	3	3	4	3.33	4	4	5	4.30	1.00
2	Match Between the System and the Real World	4	4	4	4.00	5	5	5	5.00	1.00
3	User Control and Freedom	2	4	4	3.33	5	5	4	4.67	1.33
4	Consistency and Standards	3	4	5	4.00	4	5	5	4.67	0.67
5	Error Prevention	2	4	4	3.33	4	3	4	3.67	0.33
6	Recognition Rather than Recall	4	3	5	4.00	5	4	5	4.67	0.67
7	Flexibility and Efficiency of Use	4	4	5	4.33	5	5	4	4.67	0.33
8	Aesthetic and Minimalist Design	5	4	5	4.67	4	5	4	4.33	0.33
9	Help Users Recognize, Diagnose, and Recover from Errors	3	3	4	3.33	4	3	4	3.67	0.33
10	Help and Documentation	4	4	5	4.33	5	5	5	5.00	0.67

4.2.2 Participants

Our research group has an ongoing collaboration with a local animation company, Animoka Studios (Studios 2025). The studio suggested two freelance professional storyboard artists (E1-E2), with more than 5 years of experience in 3D animation productions, to be involved in the user study.

4.2.3 The selected use case

The use case consists of creating a storyboard using the script detailed in Sect. 4.1.3 as a reference for testing all the application's functionalities. The script was available for inspection anytime in the AR/VR UIs.

4.2.4 The test procedure

The overall test procedure can be summarized as follows:

- Introduction: the user is introduced to the proposed research, and the test procedure is explained in detail
- User data: the user fills out a preliminary form used to collect general information about participants
- Test: the user is introduced to the application by a tutorial, then creates a storyboard based on the proposed use case.
- Feedback: the user provides feedback regarding the overall experience.

4.2.5 Results

The storyboard artist concluded that, based on their experience, the proposed system can effectively support storyboard artists and other experts involved in the filmmaking and animation pipelines. One of the main strengths of the proposed XR solutions is being able to freely explore the 3D space, trying out different camera shots, which allows the user to explore multiple ways to represent the scene and different storytelling options, regardless of the characters.

> The advantages of the proposed 3D tool are the possibility of recreating the 3D space, testing different framings, and navigating the space immersively; this expands the possibilities for scene representation and is particularly valuable from a creative standpoint. (E1)

On the contrary, exploring the 3D environment through desktop applications usually limits this experience.

> Desktop-based usage limits environment exploration; interactions based on panning or rotating via traditional desktop commands constrain freedom of exploration in the 3D environment and creativity. (E1)

This confirms previous studies known in the literature, as discussed in Sect. 2. Moreover, the capabilities offered by the proposed system make it a flexible tool that could benefit not only the storyboarding phase but also other steps of the production pipeline.

> I would consider the animation component to define the character poses very useful for live-action productions because it allows previewing of the scene, exploring different framings, and, if needed, producing an animatic with character animations (even simplified ones), resulting in significant time and cost savings compared to production with real locations and actors. (E1)

> Within the production pipeline, it would allow for hybridization of the storyboard phase and the 3D scene creation phase carried out by the layout artist, making it possible to clearly define the dimensions of environments, objects, and characters in the scene, particularly in complex environments or action sequences. (E2)

Moreover, the two professionals stated that the main limitation of a traditional 2D approach is

> the difficulty in correctly defining the proportions between objects and between objects and the environment; this is even more evident in complex environments. (E1)

On the contrary, the XR applications solve this limitation and allow the artist to focus on the creative process instead. Finally, both agreed that setting up the environment and freely exploring the scene immersively can greatly support the artist's creativity. The full transcript of the interviews is provided as supplementary materials.

4.3 Comparative evaluation

A priori power analysis has been performed with G*Power (Faul et al. 2009) to determine the number of participants needed to observe a statistical difference for an independent two-sample t-test. Based on preliminary tests conducted with the AR and VR applications, a strong effect size is expected. For an effect size $d = 0.8$, $alpha = 0.05$, and $Power = 0.8$, a total sample size of 42 users is expected. Due to the exploratory nature of the proposed study, a $Power = 0.8$ is usually considered reasonable in this kind of research (Giner-Sorolla et al. 2024). Thus, a *between-subjects* evaluation has been performed to assess storyboard artists' preferences and to answer the research questions detailed in Sect. 1 Two groups of twenty-one users each were asked to perform a brief storyboard using either the AR or VR interface. With regards to the proposed workflow depicted in Fig. 3, only the second step was tested by the users. Figure 5 shows one user storyboarding with the AR applications (on the left) and the VR application (on the right).

4.3.1 Metrics

To evaluate the user experience, users were asked to fill out the System Usability Scale (SUS) (Brooke et al. 1996) and the NASA Task Load Index (NASA-TLX) questionnaires (Hart and Staveland 1988) at the end of the storyboarding task. Completion times were also gathered, even if the task did not have a minimum or maximum completion time, due to the creative nature of the storyboarding process.

Errors were represented by user interactions that led to a mistake in the final storyboard: examples are users forgetting to take a screenshot, resulting in a missing panel in the final storyboard, or panels missing some actions in the descriptions.

Additional data were collected to determine users' information that can be useful to further cluster the results: age, gender, previous knowledge of the AR/VR domain, frequency of use of AR/VR devices, the most used AR/VR device, storyboard creation frequency, and preferred storyboarding tool (if any). Finally, four open questions further allowed users to openly discuss which features were most or least appreciated, as well as missing features or possible changes to existing ones.

The resulting storyboards have been visually compared to ensure that the two solutions can produce similar results. However, the comprehensibility and communication power of the resulting storyboards have not been considered in the evaluation process. The scope of the proposed research is to evaluate whether users have preferences for one of the two XR solutions. In contrast, the user's satisfaction with the obtained storyboard is already part of the SUS and NASA-TLX analysis and can be further discussed by the users in the open questions.

4.3.2 Participants

Forty-two students participated in the user tests. All the students involved in the test are enrolled either in the master's degree program in Computer Engineering (Graphic and Multimedia learning path) or the master's degree program in Cinema and Media Engineering. They also had the frequency of at least one course on 3D animation and storytelling techniques. Therefore, all the participants knew the storyboarding process well.

4.3.3 The selected use case

The use case consists of creating a storyboard using the script detailed in Sect. 4.1.3 as a reference for testing all the application's functionalities. The script was available for inspection anytime in the AR/VR UIs.

4.3.4 The test procedure

The overall test procedure can be summarized as follows:

- Introduction: the user is introduced to the proposed research, and the test procedure is explained in detail;
- User data: the user fills out a preliminary form used to collect general information about participants;
- Tutorial: the user is introduced to the application by a tutorial that describes the main functionalities of the application and lets the user practice with the interface;
- Storyboard: the user performs the use case using either the AR or the VR application;
- Questionnaires: the user has to fill out the SUS and the NASA-TLX questionnaires;

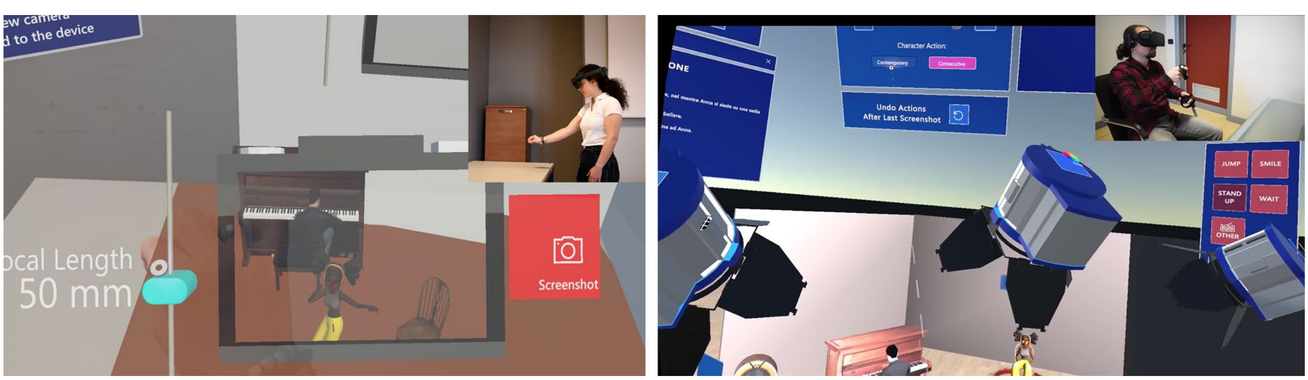

Fig. 5 A user storyboarding with the AR application (on the left) and the VR application (on the right)

- Open questions: the user can provide feedback regarding the overall experience, answering four open questions.

4.3.5 Results

All the users involved in the tests claimed to have previous knowledge of both storyboarding and the AR/VR domains. The gender distribution is similar, with 14 males and seven females in the AR test group (G_{AR}) and 13 males and eight females in the VR test group (G_{VR}). The average age is 24.62 ($\sigma = 1.62$) for G_{AR} and 23.38 ($\sigma = 2.88$) for G_{VR}. Table 3 details the user distributions in the two test groups: how often the users in the two test groups create storyboards, the preferred tools for performing this task (users can specify multiple options), how often the users in the two test groups use AR or VR applications, and their most used AR or VR device. The frequency in creating storyboards was evaluated on a (1–4) scale with $1 = rarely$ and $4 = very\_often$: the users from both groups showed a frequency in creating storyboards over average (G_{AR} $\mu = 2.8$ and $\sigma = 0.6$, G_{VR} $\mu = 2.57$ and $\sigma = 0.85$). Half of the participants in both test groups preferred paper and pen storyboarding, whereas the other half opted for digital solutions (either digital drawing or 2D/3D storyboard applications). The frequency of use of AR or VR technologies was evaluated on a (1–6) scale with $1 = never$ and $6 = everyday$: th e users from both groups showed a limited frequency of use (G_{AR} $\mu = 2.1$ and $\sigma = 1.37$, G_{VR} $\mu = 1.9$ and $\sigma = 1.04$).

Data distributions (DDs) have been examined through the Shapiro-Wilk test to assess the SUS and the NASA-TLX results. The variance for each sample has been computed, and if the ratio between the two variances was lower than 4 : 1, equal variance has been assumed for the two populations. The SUS results for G_{AR} and G_{VR} present uniform distributions ($p_{G_{AR}} = 0.586$, $W_{G_{AR}} = 0.9633$, $p_{G_{VR}} = 0.9251$, $W_{G_{VR}} = 0.98$) and the observed effect size is minimal ($KS - D_{G_{AR}} = 0.09525$, $KS - D_{G_{VR}} = 0.09783$), indicating that the magnitude of the difference between the sample distribution and the normal distributions is very small. The NASA-TLX results for G_{AR} and G_{VR} present uniform distributions ($p_{G_{AR}} = 0.3249$, $W_{G_{AR}} = 0.9489$, $p_{G_{VR}} = 0.6397$, $W_{G_{VR}} = 0.9658$) and the observed effect size is small ($KS - D_{G_{AR}} = 0.115$, $KS - D_{G_{VR}} = 0.113$), indicating that the magnitude of the difference between the sample distribution and the normal distributions is small.

Subsequently, for normally distributed data, an independent two-sample t-test was employed. The average SUS scores are $\mu_{G_{AR}} = 56.55$ ($\sigma = 19.65$) and $\mu_{G_{VR}} = 83.04$ ($\sigma = 8.44$), and the independent two-sample t-test shows statistically significant differences between the two groups ($p = 0.0000068$). The average NASA-TLX scores are $\mu_{G_{AR}} = 52.7$ ($\sigma = 20.32$) and $\mu_{G_{VR}} = 25.75$ ($\sigma = 12.32$); the independent two-sample t-test shows statistically significant differences between the two groups ($p = 0.0000072$).

The NASA-TLX outcomes have also been evaluated by comparing the six weighted dimensions. However, the sign test has been used since the data present a non-uniform distribution, and the DDs are not symmetrical. The average NASA-TLX dimension scores are depicted in Fig. 6. There are significant differences for the effort ($p = 0.04953$), frustration level($p = 0.00006$), performance (0.02535), and temporal demand ($p = 0.00729$). In contrast, it was not possible to determine statistically significant differences for the mental and physical demands ($p = 0.81855$ and $p = 0.40538$, respectively).

On average, G_{AR} users completed the task in 12 m 40 s ($\sigma = 6m20s$), whereas G_{VR} users completed the task in 14 m 29 s ($\sigma = 6m24s$). Users from G_{AR} made on average $1,58$ errors ($\sigma = 1,04$), whereas users from G_{VR} made on average $0,53$ errors ($\sigma = 0,68$). Resulting storyboards are made, on average, of 5 to 6 panels.

Table 3 How often the users in the two test groups actually create storyboards, the preferred tools used to perform this task, how often the users in the two test groups use AR (G_{AR}) or VR (G_{VR}) applications, and their most used AR/VR device

		AR	VR
Storyboarding	Very often	1	3
	Often	15	8
	Sometimes	3	8
	Rarely	2	2
Preferred tool	Paper & pen	10	11
	Digital drawing	3	5
	2D apps	6	3
	3D apps	2	4
AR/VR usage	Everyday	2	0
	More than once a week	0	1
	Once a week	1	1
	Once a month	1	2
	Less than once a month	9	9
	Never	8	8
AR/VR device	Smartphone	10	8
	MetaQuest 2	0	5
	Hololens 2	3	0

4.4 Protocol

Before tests, participants gave their written informed consent to participate, and they could discontinue participation at any time during the experiments. The experimental evaluation protocols followed the declaration of Helsinki (World Medical Association 2024). Data were recorded per the European General Data Protection Regulation 2016/679, stored in local repositories with anonymized identities, and used only for the post-processing aggregate evaluation procedure. No sensitive data was recorded. All users were

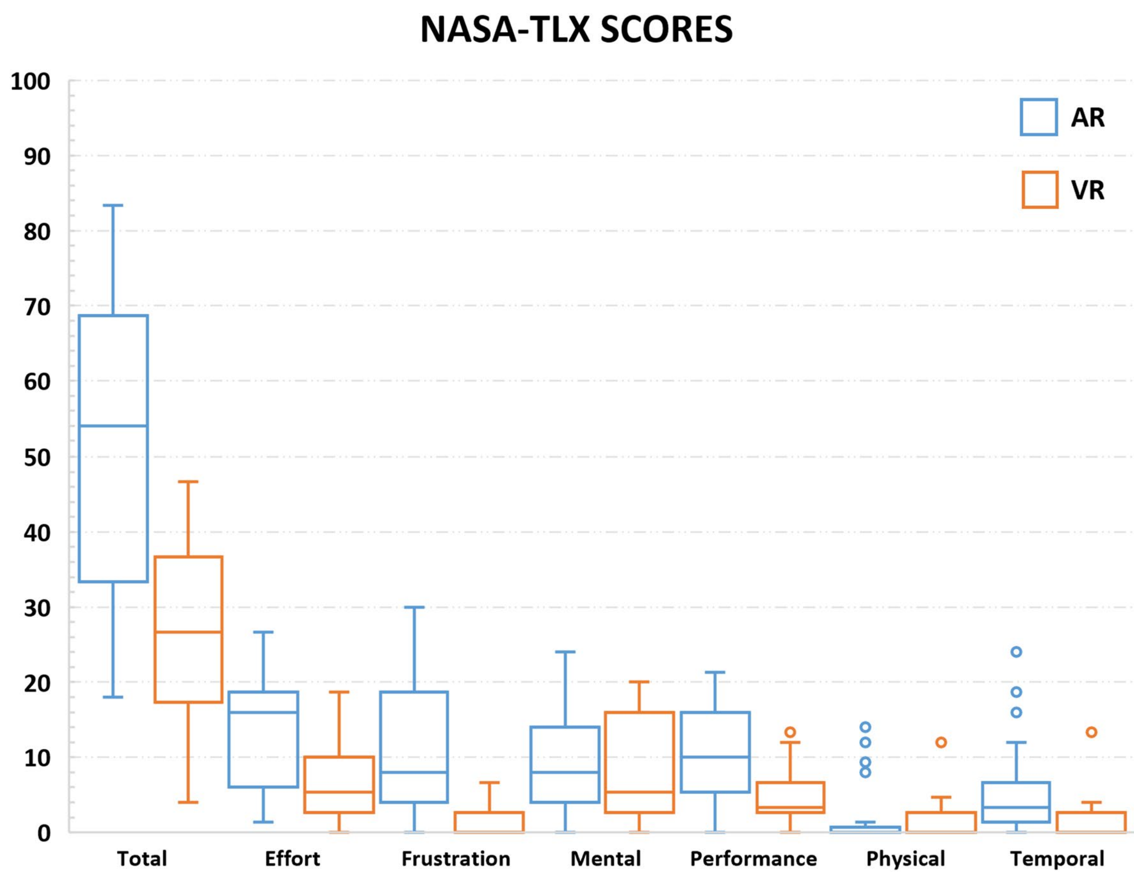

Fig. 6 The NASA-TLX Scores

volunteers, and they did not receive any reward for participating in the tests. The study was reviewed by the Politecnico di Torino's Institutional Research Ethics Committee and approved with protocol number 2024-POTOCLE-0048836.

5 Discussion

The results show that the two test groups are comparable in age, gender, storyboard expertise, and knowledge of XR technologies. Based on the NASA-TLX questionnaire's results and their statistical significance, it is possible to answer to **R3** (*what is the level of workload perceived by the users with the two solutions?*): the users perceived a higher level of workload with the AR application, with an average value ($\mu_{G_{AR}} = 52.7$) that is more than double compared to the VR one ($\mu_{G_{VR}} = 25.75$). Further analyzing each dimension individually, it seems that the Effort, Frustration, and Performance dimensions greatly affected the overall results. The VR solution allowed the users to obtain better results in the storyboarding task with less effort, as denoted by a higher average performance value and a lower average effort value compared to the AR application. The frustration level is also noticeably different in favor of the VR solution. It was not possible to assess a statistical difference pertaining to the physical and temporal domains. The physical workload is perceived as very low, with a slightly higher average value for the VR solution, which can be due to the usage of the controllers. The temporal demand is relatively low, with a higher average value for the AR application. Since the task does not imply a minimum or maximum execution time due to its creative nature, it is reasonable to assume that the higher value for the AR solution is related to its lower usability score assessed with the SUS questionnaire.

Based on the SUS questionnaire's results and their statistical significance it is possible to answer to **R2** (*what is the usability perceived by the users with the two solutions?*): the users appreciated the VR system, with an average SUS score of $\mu_{G_{VR}} = 83.04$ ($\sigma = 8.44$) that correspond to a B grade or an almost excellent rating according to the adjective rating scale proposed by Berford et al. (2017). The AR solution instead was not really appreciated by the users; the average SUS score of $\mu_{G_{AR}} = 56.55$ ($\sigma = 19.65$) corresponds to an F grade or a slightly over OK rating.

Thus, based on this data, the answer to **R1a** (*do users prefer augmented or virtual reality for creating 3D storyboards in a tabletop scenario?*) is that the users preferred the virtual reality system over the augmented reality one for creating 3D tabletop storytelling with the current interaction standards imposed by the chosen systems. To answer **R1b** (*why do they prefer one or the other application?*), it is relevant to consider both the NASA-TLX questionnaire's results and the answers provided by the users to the open questions. These optional written comments allowed the users to answer four questions:

1. Which features of the application did you like the most?
2. What did you like least or have difficulty doing?
3. What features would you add?
4. What would you change about the application?

Users from the G_{VR} greatly appreciated the proposed posing approach (9 out of 21) and the camera system (8 out of 21). The light system was also appreciated (5 out of 21). The interaction system based on the VR HMD and the controllers was considered the most relevant feature only for three users out of 21. On the other hand, no user of the G_{VR}, not even the least used with this kind of technology, reported or experienced problems related to the user interface. These results, combined with the high SUS score and the low NASA-TLX score, may suggest that the proposed VR solution was so effective that the users forgot about the physical interaction device and could focus solely on the task.

Most of the users from the G_{AR} had difficulty with the interaction system due to a perceived low tracking precision of the hand (12 out of 21) or with the voice recognition system (2 out of 21). These answers clarify the higher effort and frustration scores and the lower performance scores. It is relevant to report that only one user opted for head tracking and voice recognition to complete the task. The other users only tried it during the interactive tutorial: eventually, they performed a couple of interactions while creating the storyboard, but they mainly decided that the hand gesture approach was more natural for them. Six users pointed out that the interaction system was the best feature of the AR application. The average usage frequency of AR devices for these users was aligned with the general results (G_{AR} $\mu = 2.5$ and $\sigma = 1.5$). Thus, it is not possible to affirm that those who appreciate the interaction system the most are more expert in AR. The most appreciated feature by 13 out of 21 users from the G_{AR} was the camera system, followed by the UI flexibility (6 out of 21). The UI flexibility had a positive effect against the limited FoV of the HoloLens 2, since only one user indicated it as the most relevant limitation of the system. In contrast, two users felt overwhelmed by the number of panels in the UI. Therefore, the chosen tabletop scale was a successful strategy to mitigate the narrow HoloLens FoV. On the other hand, it is not possible to point out a relevant problem in the VR storyboarding application: negative features were never mentioned by more than three users.

5.1 Limitations

The proposed research presents some limitations. The first one is that the results cannot be generalized to the whole AR and VR categories, but only to other XR systems that provide the same visualization and interaction approach as those used in this research. This choice reflects the will to compare the storyboarding task on commercially available solutions instead of opting for an ad-hoc lab setup. To limit the comparison to the AR and VR paradigm, it would have been necessary to integrate physical controllers into the AR system: this would require a fixed external tracking system, whereas the benefit of AR mostly relies on the inside-out tracking approach. Moreover, to obtain the same FoV of the AR system on the VR system, it would have been necessary to virtually reduce the FoV on the VR application. However, whereas the reduced FoV on the AR system is balanced by the user's FoV of the real world in optical see-through solutions, this is not possible in VR without reproducing both a complete virtual environment to mimic the real world, and then adding the tabletop storyboarding visualization with a limited FoV. Instead, we opted to investigate how different standard XR commercially available systems compare using the dominant interaction strategies for the given system to perform the storyboarding task.

The second limitation is that a tabletop scale has been adopted in both applications to mitigate the problems related to the narrow FoV of AR devices. The scale factor is not an issue in a virtual environment, but very different results might be collected if an absolute dimension scale was considered in AR. The hardware selected for the VR application is not the best on the market in terms of display specifications, whereas devices such as the HTC VIVE offer better graphical quality despite a more complex setup of the working area. However, this choice did not negatively affect the test outcomes, as the users clearly preferred the VR solution. A better VR device might only lead to greater SUS scores or lower NASA-TLX scores.

6 Conclusion and future works

The proposed research investigated and compared two 3D tabletop storyboarding applications based on extended reality technologies. A scene-enacting approach with a point-and-click interaction paradigm has been adopted to define the characters' poses for the storyboard panels. The first evaluation was conducted with usability experts to identify and fix the main usability issues. Then, professional storyboard artists were involved in the user study; they appreciated the proposed system and, above all, the possibility to freely explore the scene immersively, with respect to traditional tools. Finally, a between-subjects evaluation was conducted with students to determine if they preferred one of the two proposed solutions, assessing them using the SUS and NASA-TLX questionnaires. Experimental results clearly indicate that the commercially available virtual reality-based solution outperforms the augmented reality one from the usability point of view. Moreover, a lower workload is experienced using an input interface based on physical controllers compared to the gesture-based one.

Future works will be aimed at addressing current limitations. First, it would be relevant to integrate a physical input interface, such as hand controllers or finger-tracking gloves, and evaluate how much it affects the system's usability. Another development involves overcoming the tabletop scale limitation and offering a 1:1 scale environment. In the AR scenario, an interesting option will be to use real-world elements as part of the virtual stage; integrating an object-detection and pose-estimation system may lead to a novel storyboarding paradigm, and it could partially mitigate the problems related to the limited FoV. However, testing this scenario with a passthrough AR HMD such as the Meta Quest 3 would also be interesting. This could be less impactful in the VR scenario, but it would be interesting to replace the Oculus Rift with the HTC Vive Pro, improving the freedom of movement for the user in the virtual environment and the overall experience's quality from a graphical point of view. Finally, the two professional storyboard artists suggested implementing railway paths for camera movements and a panel showing all the available camera shots of the scene concurrently.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10055-026-01381-8.

Author contributions A. S., F. M., and F. D. P. contributed to the study's conception and design. The system development and evaluation were performed by F. M., V. C., and S. P. Data analysis was performed by A. S. and F. D. P. All authors contributed to the writing and approved the final manuscript.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Data availability Data is provided within the manuscript or supplementary information files.

Code availability The code of the two applications compared in this article is publicly available on GitHub at the following link https://github.com/fmpolito/Storyboard-XR.git.

Declarations

Conflict of interest We have no conflict of interest to disclose.

Ethical approval The study was approved by the Ethical Review Board of Politecnico di Torino University (#2024-POTOCLE-0048836).

Informed consent Participants signed an informed consent form before participating in the experiment.

References

Adobe. Mixamo homepage. Accessed 25 July 2024. https://www.mixamo.com/

Adobe. Photoshop homepage. Accessed 10 September 2025. https://www.adobe.com/products/photoshop.html

Berford B, Diaz-Padron C, Kaleas T, Oz I, Penney D (2017) Building an animation pipeline for VR stories. In: ACM SIGGRAPH 2017 Talks. SIGGRAPH '17. New York, NY, USA: Association for computing machinery. Available from: https://doi.org/10.1145/3084363.3085092

Bessmeltsev M, Vining N, Sheffer A (2016) Gesture3D: posing 3D characters via gesture drawings. ACM Trans Graph. 35(6). https://doi.org/10.1145/2980179.2980240

Brooke J et al (1996) SUS-A quick and dirty usability scale. Usability Eval Ind 189(194):4–7. https://doi.org/10.1201/9781498710411-35

Canva. Canva homepage. Accessed 10 September 2025. https://www.canva.com

Chapman D (2023) Framestore Farsight – A scalable virtual production ecosystem for faster iteration. In: ACM SIGGRAPH 2023 Talks. SIGGRAPH '23. New York, NY, USA: Association for computing machinery. Available from: https://doi.org/10.1145/3587421.3595458

Clever Prototypes, LLC. Storyboardthat homepage. Accessed 10 September 2025. https://www.storyboardthat.com

Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behav Res Methods 41(4):1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Favreau L, Reveret L, Depraz C, Cani MP (2006) Animal gaits from video: Comparative studies. Graphical Models. 68(2):212–234. Special Issue on SCA 2004. https://doi.org/10.1016/j.gmod.2005.04.002

Galvane Q, Lin IS, Argelaguet F, Li TY, Christie M (2019) VR as a Content Creation Tool for Movie Previsualisation. In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR); 303–311. Available from: https://doi.org/10.1109/VR.2019.8798181

Giner-Sorolla R, Montoya AK, Reifman A, Carpenter T, Lewis NA, J, Aberson CL, et al (2024) Power to detect what? considerations for planning and evaluating sample size. Pers Soc Psychol Rev 28(3):276–301. https://doi.org/10.1177/10888683241228328
Glauser O, Ma WC, Panozzo D, Jacobson A, Hilliges O, Sorkine-Hornung O (2016) Rig animation with a tangible and modular input device. ACM Trans Graph. 35(4). https://doi.org/10.1145/2897824.2925909
Gouvatsos A. (2018) 3D storyboarding for modern animation. Bournemouth University. Available from: https://eprints.bournemouth.ac.uk/31265/
Gouvatsos A, Xiao Z, Marsden N, Zhang JJ (2014) Automatic 3D posing from 2D Hand-Drawn sketches. In: PG (Short Papers)
Gouvatsos A, Xiao Z, Marsden N, Zhang JJ (2017) Posing 3D models from drawings. Comput Entertain. 15(2). https://doi.org/10.1145/2729984
Guay M, Cani MP, Ronfard R (2013) The line of action: an intuitive interface for expressive character posing. ACM Trans Graph. 32(6). https://doi.org/10.1145/2508363.2508397
Hahn F, Mutzel F, Coros S, Thomaszewski B, Nitti M, Gross M, et al (2015) Sketch abstractions for character posing. In: Proceedings of the 14th ACM SIGGRAPH / eurographics symposium on computer animation. SCA '15. New York, NY, USA: Association for computing machinery; 185–191. Available from: https://doi.org/10.1145/2786784.2786785
Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds). Human mental workload. vol. 52 of Advances in psychology. North-Holland p. 139–183. Available from: https://doi.org/10.1016/S0166-4115(08)62386-9
Henrikson R, Araujo B, Chevalier F, Singh K, Balakrishnan R (2016) Multi-Device Storyboards for Cinematic Narratives in VR. In: Proceedings of the 29th annual symposium on user interface software and technology. UIST '16. New York, NY, USA: Association for computing machinery; 787–796. Available from: https://doi.org/10.1145/2984511.2984539
Innovative Software. FrameForge Storyboard Studio. Accessed 25 July 2024. https://www.frameforge.com/
Ishii H, Ullmer B (1997) Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of the ACM SIGCHI conference on human factors in computing systems. CHI '97. New York, NY, USA: Association for computing machinery; 234–241. Available from: https://doi.org/10.1145/258549.258715
Jacob RJK, Sibert LE (1992) The perceptual structure of multidimensional input device selection. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI '92. New York, NY, USA: Association for computing machinery. 211–218. Available from https://doi.org/10.1145/142750.142792
Jacobson A, Panozzo D, Glauser O, Pradalier C, Hilliges O, Sorkine-Hornung O (2014) Tangible and modular input device for character articulation. ACM Trans Graph. 33(4). https://doi.org/10.1145/2601097.2601112
Jain E, Sheikh Y, Hodgins J (2009) Leveraging the talent of hand animators to create three-dimensional animation. In: Proceedings of the 2009 ACM SIGGRAPH/eurographics symposium on computer animation. SCA '09. New York, NY, USA: Association for computing machinery. 93–102. Available from: https://doi.org/10.1145/1599470.1599483
Kato H, Billinghurst M (1999) Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: Proceedings 2nd IEEE and ACM international workshop on augmented reality (IWAR'99). IEEE p. 85–94. Available from: https://doi.org/10.1109/IWAR.1999.803809
Khronos Group.: Open XR. Accessed 25 July 2024. https://www.khronos.org/api/index_2017/OpenXR
Kilkenny T (2023) Virtual Cinematography with Unreal Engine: Lean how Unreal Engine's virtual cinematography framework has given rise to new and advanced techniques in film and animation studios. In: ACM SIGGRAPH 2023 Talks. 1–2
Kim H, Ali G, Hwang JI (2021) ASAP: Auto-generating storyboard and previz with virtual humans. In: 2021 IEEE International symposium on mixed and augmented reality adjunct (ISMAR-Adjunct); p. 316–320. Available from: https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00071
Lin A, Lee GS, Longson J, Steele J, Goldberg E, Stefanovic R (2015) Achieving real-time playback with production rigs. In: ACM SIGGRAPH 2015 Talks. Association for Computing Machinery; 1–1
Louarn A, Christie M, Lamarche F (2018) Automated staging for virtual cinematography. In: Proceedings of the 11th ACM SIGGRAPH conference on motion, interaction and games. MIG '18. New York, NY, USA: Association for computing machinery; Available from: https://doi.org/10.1145/3274247.3274500
Mahmudi M, Harish P, Le Callennec B, Boulic R (2016) Artist-oriented 3D character posing from 2D strokes. Comput Graph. 57(C):81–91. https://doi.org/10.1016/j.cag.2016.03.008
Manuri F, Mafrici F, Sanna A (2025) Storyboarding in extended reality: Leveraging real-world elements in storyboard creation. In: MIG '25: Proceedings of the 17th ACM SIGGRAPH conference on motion, interaction, and games. association for computing machinery. Available from: https://doi.org/10.1145/3769047.3769062
Manuri F, Sanna A, De Pace F (2023) Storytelling in the metaverse: From desktop to immersive virtual reality storyboarding. In: 2023 IEEE international conference on metrology for extended reality, Artificial intelligence and neural engineering (MetroXRAINE); 28–33. Available from: https://doi.org/10.1109/MetroXRAINE58569.2023.10405763
Manuri F, Sanna A, Scarzello M, De Pace F (2024) A novel approach to 3D storyboarding. In: Clayton M, Passacantando M, Sanguineti M, (eds) Intelligent technologies for interactive entertainment. Cham: Springer Nature Switzerland. 178–192. Available from: https://doi.org/10.1007/978-3-031-55722-4_13
Mao C, Qin SF, Wright DK (2005) A sketch-based gesture interface for rough 3D stick figure animation. In: SBM. Available from: https://doi.org/10.2312/SBM/SBM05/175-183
Microsoft. Mixed Reality Toolkit 2. Accessed 25 July 2024. https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05
Moran G (2015) Pushing photorealism in "a boy and his kite". In: ACM SIGGRAPH 2015 computer animation festival. Association for computing machinery. 190–190
Mouajjah S, Plessiet C (2021) 3D character posing from 2D sketches for storyboarding. In: 2021 7th international HCI and UX conference in Indonesia (CHIuXiD). 1, 12–15. Available from: https://doi.org/10.1109/CHIuXiD54398.2021.9650611
Nebeling M, Rajaram S, Wu L, Cheng Y, Herskovitz J (2021) XRStudio: A virtual production and live streaming system for immersive instructional experiences. In: Proceedings of the 2021 CHI conference on human factors in computing systems. CHI '21. New York, NY, USA: Association for computing machinery. Available from: https://doi.org/10.1145/3411764.3445323
Nielsen J (1994) Enhancing the explanatory power of usability heuristics. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI '94. New York, NY, USA: Association for computing machinery; 152–158. Available from: https://doi.org/10.1145/191666.191729
Nielsen J, Molich R (1990) Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI '90. New York, NY, USA: Association for

computing machinery; 249–256. Available from: https://doi.org/10.1145/97243.97281

Pizzi D, Lugrin JL, Whittaker A, Cavazza M (2010) Automatic generation of game level solutions as storyboards. IEEE Trans Comput Intell AI Games 2(3):149–161. https://doi.org/10.1109/TCIAIG.2010.2070066

Rao A, Jiang X, Guo Y, Xu L, Yang L, Jin L, et al (2023) Dynamic storyboard generation in an engine-based virtual environment for video production. In: ACM SIGGRAPH 2023 Posters. SIGGRAPH '23. New York, NY, USA: Association for Computing Machinery; Available from: https://doi.org/10.1145/3588028.3603647

Schaefer M (2024) Designing in VR. Accessed 25 July 2024. http://www.mattschaeferdesign.com/designing-in-vr

Shin M, Kim Bs, Park J (2005) AR storyboard: an augmented reality based interactive storyboard authoring tool. In: Fourth IEEE and ACM international symposium on mixed and augmented reality (ISMAR'05). IEEE, 198–199. Available from: https://doi.org/10.1109/ISMAR.2005.12

Siddi F. Blender For Pre-Production. Accessed 10 September 2025. https://www.blender.org/user-stories/the-future-of-storyboarding-blender-for-pre-production/

Sito T (2013) Moving Innovation: A History of Computer Animation. The MIT Press; Available from: http://www.jstor.org/stable/j.ctt5vjqq9

Studios A.: Animoka Studios homepage. Accessed 10 September 2025. https://www.animoka.com/

Trottnow J, Götz K, Seibert S, Spielmann S, Helzle V, Einabadi F, et al (2015) Intuitive virtual production tools for set and light editing. In: Proceedings of the 12th European conference on visual media production. CVMP '15. New York, NY, USA: Association for computing machinery. Available from: https://doi.org/10.1145/2824840.2824851

Truong KN, Hayes GR, Abowd GD (2006) Storyboarding: an empirical determination of best practices and effective guidelines. In: Proceedings of the 6th conference on Designing Interactive systems, p. 12–21

Unity Technologies. Unity 3D. Accessed 25 July 2024. https://unity.com/

WebGames3D com. Shot Professional. Accessed 25 July 2024. https://www.shotprofessional.com/

Wei X, Chai J (2011) Intuitive interactive human-character posing with millions of example poses. IEEE Comput Graphics Appl 31(4):78–88. https://doi.org/10.1109/MCG.2009.132

World medical association: Declaration of Helsinki. Accessed 25 July 2024. https://tinyurl.com/7s6djxpn

Xiao J, Tang Z, Feng Y, Xiao Z (2015) Sketch-based human motion retrieval via selected 2D geometric posture descriptor. Signal Process 113:1–8. https://doi.org/10.1016/j.sigpro.2015.01.004

Xu K, Chen K, Fu H, Sun WL, Hu SM (2013) Sketch2Scene: sketch-based co-retrieval and co-placement of 3D models. ACM Trans Graph. 32(4). https://doi.org/10.1145/2461912.2461968

Zhang JQ, Xu X, Shen ZM, Huang Z, Zhao Y, Cao YP, et al (2021) Write-an-animation: High-level text-based animation editing with character-scene interaction. Computer Graphics Forum. 40. https://doi.org/10.1111/cgf.14415