Volumivive: An Authoring System for Adding Interactivity to Volumetric Video

Qiao Jin[1] Yu Liu[1] Puqi Zhou[2] Bo Han[2] Svetlana Yarosh[1] Feng Qian[1]*

[1]University of Minnesota [2]George Mason University

ABSTRACT

Volumetric video is a medium that captures the three-dimensional (3D) shape and movement of real-life objects or people. However, pre-recorded volumetric video is limited in terms of interactivity. We introduce a novel authoring system called Volumivive, which enables the creation of interactive experiences using volumetric video, enhancing the dynamic capabilities and interactivity of the medium. We provide four interaction methods that allow users to manipulate and engage with digital objects within the volumetric video. These interactive experiences can be used in both augmented reality (AR) and virtual reality (VR) settings, providing users with a more immersive and interactive experience.

1 INTRODUCTION

Volumetric video is a video technique that records real-life three-dimensional (3D) shape and movement of objects or people. This allows for a more realistic and lifelike representation of the captured subjects, as opposed to traditional video which only captures the two-dimensional image of a scene, only showing the subject from a fixed perspective. Volumetric video brings new opportunities to create new types of content that were previously not possible, such as realistic and immersive experiences that allow users to view objects or environments from different angles. For example, volumetric video could be used to create a virtual tour of a museum exhibit, allowing users to walk around and view artifacts from different angles and distances. It could also be used to create training simulations that allow users to practice skills in a natural and engaging way. Additionally, volumetric video has the potential to be used in a variety of other applications, such as film, television, and gaming, to create more authentic and immersive experiences for audiences. Fields of architecture, engineering, and design can also use volumetric videos to create more accurate and detailed digital representations of objects and environments.

In this work, we present an authoring system called Volumivive for adding interactivity to volumetric video. Volumivive can be used to create interactive augmented reality (AR) and virtual reality (VR) experiences that allow users to manipulate and interact with digital objects and environments recorded in volumetric video, expanding the potential and range of applications for volumetric video. Our contributions include a system workflow for authoring interactive volumetric video, from video capturing to interaction authoring and video display; and a user interface and interactive assets for volumetric video editing and manipulation.

2 BACKGROUND

Many 3D virtual environments are based on textured polygonal models that represent the appearance and geometry of the virtual world. Creating a video of a virtual environment with realistic and dynamic objects usually has high requirements of 3D modeling and animation or other computer graphics techniques, which can allow for flexible control and customization of the 3D objects or avatars' appearance, movement, and response to users' input. Volumetric video differs in the level of detail and realism, as well as the techniques used to create it. Volumetric video captures the real-world shape and movement of objects or people directly using multi-view photography or depth sensing, which can result in a more realistic and lifelike representation but may be subject to limitations in customization and interactivity.

Although live volumetric video has the potential to create highly immersive and interactive augmented reality (AR) and virtual reality (VR) experiences (e.g., [3]), it has a number of challenges and limitations. One major challenge is the high computational and hardware requirements for real-time volumetric video capture and rendering. This can make it difficult to achieve the level of detail and realism desired, particularly for applications that require a large number of objects or a wide field of view. Another challenge is the need for specialized hardware and infrastructure to capture and render volumetric video in real time. It is usually costly to set up and maintain the necessary systems, which may be a barrier to adoption for some applications. Additionally, live volumetric video may require the use of markers or other tracking systems to accurately capture and render three-dimensional information about an object or scene. This can add complexity to the setup and may limit the range of motion or interaction possible. For those reasons, volumetric video is typically pre-recorded (e.g., [4,5]), which means that it does not allow for real-time interaction or manipulation with digital objects by the user. This can limit the sense of immersion and interactivity compared to other types of AR/VR experiences, which may be more responsive to user inputs.

This articulates a research gap in the development and investigation of interactivity authoring systems for pre-recorded volumetric video. In this work, we empower users to customize the way to interact with the content of volumetric video, allowing users to select different viewpoints and create and interact with interactive areas within the video.

3 SYSTEM DESIGN

Volumivive is a volumetric video authoring system consisting of three key components: capturing, authoring, and displaying. We designed our system with Unity3D and C# on Oculus Quest I.

3.1 Capturing

As there are no sufficient open-sourced volumetric video materials available currently, we seek to extend the creation and generalization scale by providing the capturing side. We used six Microsoft Kinect 2 [1] and four calibration markers for capturing volumetric videos, as shown in Figure 1a. Each Kinect camera was connected to a Windows desktop or laptop to collect the raw volumetric data and stream the data to the server desktop, where all the raw data would be calibrated and merged together to generate the final volumetric video frame. In order to maintain the best capturing quality, we

*email: fengqian@umn.edu

[1]https://learn.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows

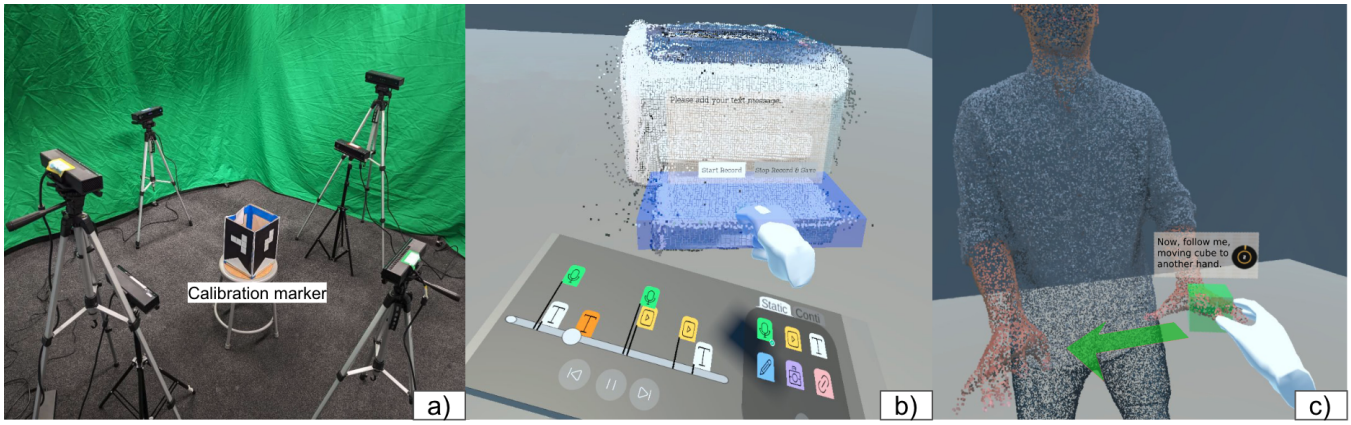

Figure 1: a) View of our capturing system with six Kinect cameras and four calibration markers; b) User was adding an interaction area with static trigger and pop-up information; c) User was interacting with the volumetric video through a continuous trigger while the video is playing.

set up a studio with green background and consistent lighting. We used LiveScan3D [2] to calibrate all the cameras and obtain the raw volumetric video data. As our design platform (Unity 3D) could not natively load this type of raw data, we used a custom converter to transform the raw data into a self-defined format that could be imported into Unity 3D.

The average point density of the video we created is around 20K points/frame, after removing the background and noise. The visual quality and video detail can be maintained and revised by this point density based on the user's requirement.

3.2 Authoring and Displaying

Volumivive allows users to create *interactive areas (IAs)* that highlight specific 3D areas (shown as cubes in Figure1b and 1c) in space and trigger specific outcomes when interacting with them. Users could add, delete, resize, and place the IA to the video content during the authoring session, then view and interact with the video during the displaying session. Each IA has three key attributes: interaction trigger, video status and interaction outcome. Interaction trigger refers to how user initiates the interaction with the IA. This includes the interaction duration, gestures, etc. Video status refers to when the user can interact with the IA. Interaction outcome defines the output of an interaction. Users can easily make a volumetric video interactive by defining the three attributes when placing an IA. Next, we explain the three attributes in detail.

Interaction Trigger. We designed two types of interaction triggers: static and continuous. *Static trigger* refers to a single interaction (e.g., touch, hit, etc) with the interaction area. Despite its simple nature, the static trigger can also support multiple interaction techniques by defining the interaction gesture or direction. For example, pushing a button from only the top part of the IA, squeezing a ball with the "grab" gesture on the IA, cutting an object from the top, etc. *Continuous trigger* refers to a series of successive interactions (e.g., writing in the air, following a hand movement) where users need to follow the IA during the interaction. The continuous trigger IA can also support flexible interaction by defining the IA's movement and interaction duration. For example, attaching a continuous IA to a moving object and following the movement can be used for learning dancing, exercising, social interactions, etc.

Video Status. Same to the interaction trigger, we defined two video statuses for IA: static and continuous. For the *Static Status*, interaction happens while the video is paused, whereas, for the *continuous status*, interaction happens while the video is still playing. Having interactions while the video is paused gives more flexibility to the interaction since there is no time limit for the interaction. However, interacting with the video while it's paused also reduces the sense of immersion. Compared to static status, continuous status provides a higher sense of immersion at the cost of lower interaction accuracy because the IA might be moving, and it's harder to interact with it accurately.

Interaction Outcome. For the current system, we considered the two most common interaction outcomes of interactive video [1]: pop-up information and branch video display. Pop-up information boxes provide extra information about the video content, including texts, links, pictures, etc. The branch display shows different video clips based on different interaction inputs. It makes the video more immersive and interactive but requires extra video clips and effort to create the video. The more video clips that are captured, the more developments and outcomes that the video can lead to. Both interaction outcomes are supported by the triggers and video status.

4 Conclusion and Future Work

We introduced a novel interaction authoring system Volumivive, seeking to extend the volumetric video with dynamic and customized interactivity. Our system can be used in the domains like advertisement, education and entertainment. We are working on completing the detailed system design and implementation. In the future, we will extend our authoring system by integrating computer vision methods to support interaction areas following the specific moving components within the video automatically. We will also conduct a preliminary user study to testify to our system's usability.

References

[1] S. Fels, E. Lee, and K. Mase. Techniques for interactive video cubism. In *Proceedings of the eighth ACM international conference on Multimedia*, pp. 368–370, 2000.

[2] M. Kowalski, J. Naruniec, and M. Daniluk. Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors. In *2015 international conference on 3D vision*, pp. 318–325. IEEE, 2015.

[3] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 741–754, 2016.

[4] N. O'Dwyer, G. W. Young, N. Johnson, E. Zerman, and A. Smolic. Mixed reality and volumetric video in cultural heritage: Expert opinions on augmented and virtual reality. In *International conference on human-computer interaction*, pp. 195–214. Springer, 2020.

[5] S. Subramanyam, J. Li, I. Viola, and P. Cesar. Comparing the quality of highly realistic digital humans in 3dof and 6dof: A volumetric video case study. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 127–136. IEEE, 2020.