A Virtual Reality Volumetric Music Video: Featuring New Pagans

Gareth W. Young*

Trinity College Dublin, Dublin, Ireland. YoungGa@tcd.ie

Abstract. Music videos are creative short films that showcase songs and visuals for both artistic and promotional purposes. With advancements in technology, modern music videos now use various techniques and post-production tools to offer diverse and engaging experiences for audiences on multiple devices. One such technology, volumetric video (VV), is gaining popularity in capturing and reproducing live musical performances in 3D as volumetric music videos (VMVs). These 3D reproductions are designed to provide cutting-edge audiovisual entertainment for both traditional 2D screens and emerging extended reality (XR) platforms like augmented and virtual reality (AR/VR). However, the impact of VMVs on virtual production techniques is still uncertain and evolving. In this short paper, we describe the creation of a unique VMV that applies VV for presentation via VR and hypothesize on what this means for future music productions and serious game technologies.

Keywords: Virtual Reality, Volumetric Video, Music Video

1 Introduction

In this work, we demonstrate a volumetric music video (VMV) presented in virtual reality (VR) that aims to explore volumetric video (VV) representations of music performances using VR technology. It serves as a sophisticated and interactive music video that can be accessed and presented through various XR platforms, showcasing new workflows for capturing, editing, and accessing VMVs for virtual production. Through this approach, we aim to gain a better understanding of how the industry can harness this technology from a serious games perspective, demonstrating how professionals are likely to react to music videos in an XR context and providing insights into virtual simulations for interactive digital media training and research.

Virtual production techniques that use XR technology can be considered serious gaming as such technology can be used to create interactive virtual environments (IVEs) for serious games. Serious gaming typically refers to games that are designed for a specific purpose beyond entertainment, such as education, training, simulation, or research [9,20]. While virtual production techniques can be used in these contexts, they require inherently designed interactions for

* Supported by V-SENSE and Volograms

them to function effectively. Rather, virtual production techniques that use XR technology are primarily used in film, television, and other forms of media production to create immersive, interactive, and dynamic environments. Still, the use of XR technology in virtual production can create opportunities for new forms of serious gaming, such as virtual simulations for training or research [1,2]. In our demonstration, we hope to show the potential of this technology in future interactive VV serious game applications.

2 Background

The rapid resurgence of VR technology [3] has led to a new quest for novel ways to visualize and communicate musical performance, driven by artistic creativity, technological innovation, and the desire to capture the attention of new and existing audiences. It has been widely acknowledged that artificial reality is the definitive technological expression of the postmodern condition; therefore, it is possible to express postmodernist art representations within VR [12, p.169].

The use of XR technology in virtual production has created opportunities for new forms of serious gaming, where studio professionals and learners can interact in real-time with the virtual world in a highly immersive and realistic way [4]. With XR technology, virtual production can simulate real-world environments and enable the player to experience them in a fully immersive way [2]. This can enhance the realism and engagement of the gameplay, allowing players to feel as though they are truly influential in the game world.

One example of this is the use of virtual production and XR technology in creating training simulations for industries such as aviation [7], healthcare [11], and the military [11]. By using VR or AR headsets, trainees can practice complex procedures and scenarios in a safe and controlled environment, without the risk of causing harm or damage. This can help to reduce training costs and improve the effectiveness of training programs [6]. Overall, the use of XR technology in virtual production can create exciting opportunities for new forms of serious gaming, as well as other forms of interactive and immersive media. By combining the power of virtual environments with real-time rendering, motion capture, and other cutting-edge technologies, virtual production using XR is poised to revolutionize the way we create and experience media.

Another example of serious gaming using XR technology is in the field of interactive digital media education [5]. VR and AR can be used to create interactive learning experiences that can engage students in a way that traditional methods cannot. For example, virtual field trips can be created that allow students to explore historic sites, and scientific phenomena, or even travel to other countries, all without leaving the classroom [19]. This can not help to make learning more engaging and accessible for students of all ages and backgrounds but can be used to educate new media practitioners about the natural affordances of immersion and presence via XR technology.

In this project, we present an immersive VMV as an emergent art form, using various XR technologies such as stereoscopic and 360-degree audiovisual spatial recording technology [13]. These capture technologies have expanded the viewing experience beyond the traditional medium and provided new dimensions of immersion, interaction, and imagination for the audience, with advancements in home PC GPU/CPU speeds, HMD optics, software data processing capabilities, and AI being closely tied to these advancements [18].

Informed by the study of XR and VMVs [15], we have designed a custom-made VV VR music video experience, featuring the New Pagans' track *Lily Yeats*, with a user-centered design approach. The project's pilot study has highlighted specific qualities that audiences seek when consuming such materials [15]. Further iterations of this application have focussed on the differences between traditional media and new XR experiences and build upon existing studies in human-computer interaction (HCI) that focus on music, haptics, and technology in use[8], specifically, those examining how users experience and potentially learn about virtual production from music videos presented via 6 DoF XR technologies [17].

3 Technical Description

VV is a technique used to capture a three-dimensional representation of a real-world object or scene, including people, objects, and environments [14]. It involves capturing a large number of images or video frames of the object or scene from multiple angles using an array of cameras or sensors, and then processing and combining the data to create a three-dimensional representation that can be viewed from any angle [10]. VV is different from traditional video in that it allows the viewer to move around and explore the object or scene in three dimensions as if they were actually there. This makes VV an ideal technology for creating immersive experiences in virtual and augmented reality, as well as for other applications such as video conferencing, gaming, and training.

One of the benefits of VV is that it allows for the creation of realistic, lifelike representations of people and objects that can be used in a variety of contexts [16]. For example, VV can be used to create virtual actors for film and video games or to create training simulations for industries such as healthcare and aviation. However, creating high-quality VV requires a significant amount of time, expertise, and resources. It typically involves using specialized equipment and software to capture and process the data and requires a team of experts to oversee the entire process. Despite these challenges, VV is an exciting and rapidly developing technology that has the potential to transform the way we interact with the world around us.

VMVs are a type of music video that uses VV technology to capture a three-dimensional representation of a live music performance. VMVs allow viewers to experience the music performance in an immersive and interactive way as if they were actually present at the performance. In a VMV, the performer or performers are recorded using an array of cameras or sensors from multiple angles, capturing a three-dimensional representation of the performance. This data is then processed and combined to create a three-dimensional video that can be viewed from any angle, providing a highly immersive experience. One notable example is the musician and producer Travis Scott, who collaborated with Fortnite, an online video game, to create a virtual concert experience called “Astronomical” in April 2020. The concert utilized volumetric technology to render Travis Scott’s performance as a 3D model and project it within the game environment, providing an immersive experience for the players.

VMVs are a relatively new technology and are still being explored and developed. However, they have the potential to revolutionize the way we experience music performances, allowing us to feel like we are truly part of the performance and providing new opportunities for artists to connect with their fans. There are many different ways that VMVs can be used, including as part of a promotional campaign for a new album or single, as a way to provide a more engaging and interactive concert experience for fans, or as a way to preserve live performances for future generations. Overall, VMVs are an exciting development in the world of music and entertainment and are likely to become more common as XR technology continues to develop and evolve.

4 New Pagans Capture Process

The VV capture process utilizes photogrammetry principles to create a 3D volumetric model with texture for each video frame. The New Pagans musicians, Lyndsey McDougall and Cahir O’Doherty were recorded separately performing in the VV capture studio using twelve video cameras. The cameras were positioned in each corner of the performance space and recorded against a greenscreen backdrop. The videos from each camera were synchronized to ensure that the captured gestures on one camera aligned with the same frame number on any other given camera. The audio was recorded separately and played back during the VV capture performance, matching the video files’ length and playback speed. This method simplified the post-production chroma-keying processes.

The process for creating VV involves recording individual camera clips of a performance, trimming and aligning them, then segmenting the central performance figure to generate a silhouette for each frame. This data is then input into a postproduction process that uses 3D reconstruction algorithms to create dynamic 3D models, which are merged using a bespoke combinational method. The resulting VV assets are imported into the Unity game engine using a custom-built SDK and played back at a standard video frame rate of 30fps to create the illusion of movement. Although VV technology has limitations, it allows for high-fidelity capture and display of musicians and enables the creation of immersive and interactive music video content. The VR experience can be accessed via OpenXR, an open standard for multiple VR platforms and devices.

A demonstration of the potential of this technology for VMV production can be seen at the following link: https://youtu.be/0Q8zUpefKt8.

Fig. 1. New Pagans performing in the capture studio

5 Advantages and Disadvantages of Using VV

VVs can be used to create immersive and realistic content, but they have distinct advantages and disadvantages. VV captures the entire 3D scene, including depth information, allowing for a highly realistic and immersive representation of the environment – this makes it well-suited for applications with crucial spatial contexts, such as architectural visualization and virtual tourism. By capturing the physical object within the main scene, VV reduces the need for extensive post-processing to animate or rig characters, saving time and resources in the production pipeline. Furthermore, unlike some motion capture (MOCAP) systems, it does not require actors or objects to be outfitted with markers or motion capture suits, reducing setup time and allowing for more natural and unencumbered performances. Finally, VVs can be interactive, allowing users to explore and interact with the captured scene from different angles. It can also accommodate dynamic scenes with moving elements, making it suitable for interactive storytelling and gaming.

Conversely, VV generates large amounts of data due to its three-dimensional nature – this requires significant processing power and storage capacity, which can increase production costs. VV may also not capture fine details as effectively as traditional MOCAP, especially for close-ups of characters or objects. MOCAP often provides more precise control over individual body parts. Capturing complex characters with intricate facial expressions and detailed movements can be challenging in VV, particularly compared to MOCAP techniques targeting detailed character animation.

Another point of contention is the uncanny valley – this refers to the phenomenon where humanoid objects or characters that closely resemble humans trigger feelings of eeriness or discomfort when they fall short of looking completely realistic. Both volumetric video and motion capture can impact the perception of the uncanny valley in different ways.

In the pursuit of addressing the uncanny valley phenomenon, the advantages of VV lie in its capacity to capture a holistic 3D environment, engendering a realistic spatial context that aids in attenuating the unsettling effect. Furthermore, its ability to capture natural movements and facilitate interactive engagement with the environment lends a sense of fluidity and immersion that mitigates potential imperfections in individual character animations. However, VV may encounter limitations in preserving intricate details, particularly concerning facial expressions and nuanced character animations, which can hinder the attainment of heightened realism. In contrast, MOCAP offers precise and detailed character animations, including subtle facial expressions, that contribute to bridging the uncanny valley gap. Additionally, MOCAP provides animators with granular control over character performances, enabling adjustments to avoid unnatural movements. Yet, this advantage is juxtaposed by the potential for artifacts or distortions resulting from markers or suits used in the process. Moreover, the restricted focus of MOCAP on movements may inadvertently detach characters from their environmental context, potentially accentuating the uncanny valley effect. Consequently, the choice between VV and MOCAP as tools for managing the uncanny valley effect hinges upon specific project demands and the intricate interplay of realism, interactivity, and contextual fidelity.

6 Conclusions

XR technology has been a game-changer for virtual production, offering a range of tools and techniques for creating immersive and interactive experiences for film, television, and gaming. In virtual production, XR can be used to create virtual environments that are indistinguishable from real-life, allowing filmmakers to create anything they can imagine. Virtual production techniques that use XR technology can indeed create opportunities for new forms of serious gaming. Whether it's in the areas of training, education, or entertainment, the use of XR technology can enhance the immersive and realistic experience of the player, leading to more effective and engaging gameplay.

7 Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 15/RP/2776 and the Horizon Europe Framework Program (HORIZON), under grant agreement 101070109. The volumetric video content for the music video was generated by Volograms (www.volograms.com). More information about New Pagans and their music can be found at www.newpagans.com.

References

1. Aldrich, C.: The complete guide to simulations and serious games: How the most valuable content will be created in the age beyond Gutenberg to Google. John Wiley & Sons (2009)

2. Checa, D., Bustillo, A.: A review of immersive virtual reality serious games to enhance learning and training. Multimedia Tools and Applications **79**, 5501–5527 (2020)
3. Evans, L.: The re-emergence of virtual reality. Routledge (2018)
4. Kavakli, M., Cremona, C.: The virtual production studio concept–an emerging game changer in filmmaking. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). pp. 29–37. IEEE (2022)
5. Meyer, O.A., Omdahl, M.K., Makransky, G.: Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. Computers & Education **140**, 103603 (2019)
6. Naranjo, J.E., Sanchez, D.G., Robalino-Lopez, A., Robalino-Lopez, P., Alarcon-Ortiz, A., Garcia, M.V.: A scoping review on virtual reality-based industrial training. Applied Sciences **10**(22), 8224 (2020)
7. Oberhauser, M., Dreyer, D.: A virtual reality flight simulator for human factors engineering. Cognition, Technology & Work **19**, 263–277 (2017)
8. Paterson, J., Wanderley, M.M.: Feeling the future—haptic audio. In: Arts. vol. 12, p. 141. MDPI (2023)
9. Susi, T., Johannesson, M., Backlund, P.: Serious games: An overview. Technical Report HS- IKI -TR-07-001 (2007)
10. Valenzise, G., Martin, A., Zerman, E., Ozcinar, C.: Immersive Video Technologies. Academic Press (2022)
11. Viglialoro, R.M., Condino, S., Turini, G., Carbone, M., Ferrari, V., Gesi, M.: Augmented reality, mixed reality, and hybrid approach in healthcare simulation: a systematic review. Applied Sciences **11**(5), 2338 (2021)
12. Woolley, B.: Virtual worlds: A journey in hype and hyperreality. Penguin, London (1993)
13. Young, G.W., O'Dwyer, N., Smolic, A.: A virtual reality volumetric music video: featuring new pagans. In: Proceedings of the 13th ACM Multimedia Systems Conference. pp. 331–333 (2022)
14. Young, G.W., O'Dwyer, N., Smolic, A.: Volumetric video as a novel medium for creative storytelling. In: Immersive video technologies, pp. 591–607. Elsevier (2023)
15. Young, G.W., O'Dwyer, N., Moynihan, M., Smolic, A.: Audience experiences of a volumetric virtual reality music video. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). pp. 775–781. IEEE (2022)
16. Young, G.W., O'Dwyer, N., Smolic, A.: Exploring virtual reality for quality immersive empathy building experiences. Behaviour & Information Technology **41**(16), 3415–3431 (2022)
17. Young, G.W., O'Dwyer, N., Vargas, M.F., Donnell, R.M., Smolic, A.: Feel the music!—audience experiences of audio–tactile feedback in a novel virtual reality volumetric music video. In: Arts. vol. 12, p. 156. MDPI (2023)
18. Young, G.W., Smolic, A.: Extended reality: music in immersive xr environments: the possibilities (and approaches) for (ai) music in immersive xr environments. In: Artificial Intelligence and Music Ecosystem, pp. 68–82. Focal Press (2022)
19. Young, G.W., Stehle, S., Walsh, B.Y., Tiri, E.: Exploring virtual reality in the higher education classroom: Using vr to build knowledge and understanding. Journal of Universal Computer Science **26**(8), 904–928 (2020)
20. Young, M.F., Slota, S., Cutter, A.B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., Yukhymenko, M.: Our princess is in another castle: A review of trends in serious gaming for education. Review of educational research **82**(1), 61–89 (2012)