

Musical Instrument Performance in Augmented Virtuality

Wanwan Li
wli17@gmu.edu
George Mason University
Fairfax, Virginia, USA



Figure 1: This figure shows a user (left) who is watching the musical instrument performance in Augmented Virtuality (AV) through VR glasses. In this case, a musician (right) is playing the cello in a virtual 3D scene.

ABSTRACT

In this paper, we present the preliminary study results about musical instrument performance in Augmented Virtuality (AV). In our proposed approach, musician can prerecord their performance using any camera device. Then, through the processing pipeline that we have proposed, the musicians are able to play musical instruments in entirely virtual worlds. According to the 3D virtual scene set up for Virtual Reality (VR) glasses, the audience can watch the performance in virtual reality where the musician is playing music in the real world. In the end, we rendered the musician's performance into 360 VR video to enhance the immersiveness of the user experiences.

CCS CONCEPTS

• **Computing methodologies** → **Rendering**; **Virtual reality**; • **Applied computing** → **Performing arts**.

KEYWORDS

computer graphics, augmented virtuality, musical instrument performance

ACM Reference Format:

Wanwan Li. 2022. Musical Instrument Performance in Augmented Virtuality. In *2022 6th International Conference on Digital Signal Processing (ICDSP) (ICDSP 2022)*, February 25–27, 2022, Chengdu, China. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3529570.3529586>

1 INTRODUCTION

With the rapid development of advanced technologies in computer graphics, switching between reality and virtuality become a very hot topic among computer graphics researchers. Especially, developing applications that are extending reality with virtuality introduce a new dimension of human-computer interactions. Nowadays, the two most popular computer graphics technologies people are talking about is Virtual Reality (VR) and Augmented Reality (AR). According to how much degree to which the reality is extended with virtuality, applications in eXtended Reality (XR) are mainly divided into four categories: Reality (R), Augmented Reality (AR), Augmented Virtuality (AV), and Virtuality (V) or Virtual Reality (VR). However, due to the limitations on computer vision technologies, Augmented Virtuality (AV) has not been widely studied by researchers so far. This is mainly because of the poor quality of the video background removal using traditional computer vision technologies. Therefore, the immersiveness in Augmented Virtuality (AV) is not comparable to Augmented Reality (AR) and Virtual Reality (VR) whose technologies are much more mature.

Figure 2 shows an example to illustrate the difference between the concepts of Augmented Reality (AR) and Augmented Virtuality (AV) in a straightforward way: A cello player is playing the cello in a room, his real performance is recorded through the camera device (b). By adding the virtual objects such as 3D houses and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICDSP 2022, February 25–27, 2022, Chengdu, China

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9580-9/22/02...\$15.00

<https://doi.org/10.1145/3529570.3529586>

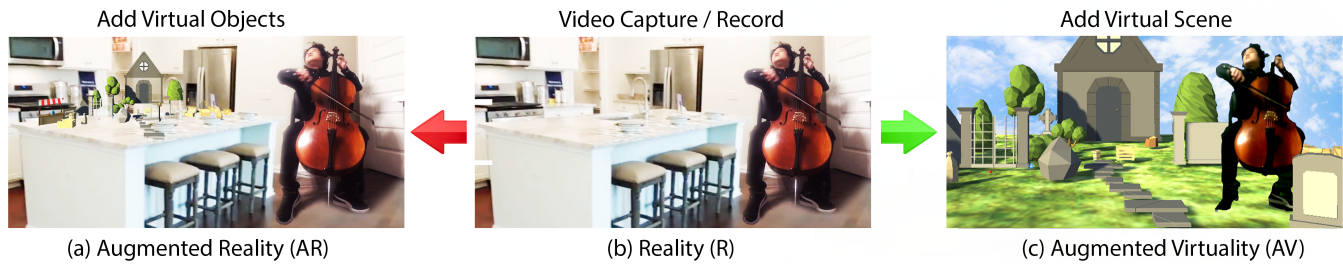


Figure 2: Introduction of Augmented Virtuality (AV).

trees in the real scene such as a table captured in the video, the video in Reality (R) is extended into Augmented Reality (AR) as shown in (a). Therefore, the red arrow points out a direction that is from Reality (R) to Augmented Reality (AR). On the other side, by replacing the real scene background with a virtual scene, in this case, it is a virtual garden. Then, the captured video in Reality (R) is extended into Augmented Virtuality (AV) as shown in (c). Therefore, the green arrow points out a direction that is from Reality (R) to Augmented Virtuality (AV). Different from the challenges in AR which is focusing on camera calibrations and depth estimation, the challenges in AV are more focused on the high-resolution real-time background removal, foreground depth estimation, and stylized/realistic rendering in the video.

Until now, there is a growing trend to use AV technologies in daily life. For example, during a virtual meeting on Zoom, which is a popular proprietary video teleconferencing software developed by Zoom Video Communications, presenters are allowed to switch the background in the webcam to a particular background image. At here, Zoom is using real-time background removal technology. However, the resulting video looks unnatural and the background removal algorithm sometimes can result in weird results such as missing hands or flickering background. Similar AV technologies are also widely used in video processing multimedia software such as Adobe Premiere and Adobe After Effects. Unfortunately, there are lots of requirements on the video background, for example, recording video before a green screen background. This requires lots of effort in video recordings by professionals. Also, it needs lots of manual efforts from the video editors to set up the threshold frame by frame to achieve the final result in background removal.

Therefore, given this open opportunity for developing an AV interface for nonprofessionals, we explore the possibilities in applying AV to musical instrument performance, which is the largest potential market for AV technology. In order to make our work easier to be adopted by more users, we make our approach more democratic by taking into account the nonprofessional video as input which requires less on the background. The contributions of our work include:

- We propose a novel topic about applying AV technology to musical instrument performance.
- We propose and implement a novel interface for musical instrument performance in AV.
- We collect the experimental results from different musicians and demonstrate our interface with different musical instruments, scenes, and settings including 360 VR video supports.

2 RELATED WORK

Augmented Virtuality (AV). At the current stage, even though there are too many limitations on existing technologies in AV, some research works still show interest in exploring how can AV be applied to innovate interfaces for human-computer interactions (HCI). In 1997, Simsarian et al. [16] proposes an example of augmented virtuality called *Windows on the world*. Later, in 2004, Regenbrecht et al. [15] proposes a novel video conferencing framework that allows users to collaborate remotely in augmented virtuality. In this interface, a virtual meeting room is built, users' videos appear on the virtual computers in the virtual meeting room. However, this work is not as immersive as our work because the user's video appears on a 2D virtual screen that is unnaturally separated from the 3D environment. At the same time, in 2005, Paul et al. [13] proposed an AV-based method on stereoscopic reconstruction in multi-modal image-guided neurosurgery. Recently, in 2010, Caballero et al. [4] proposed an AV-based interface to extend real hands with virtual world on mobile phone so that user can directly manipulate virtual three-dimensional objects inside the phone by gesturing with hands. In 2015, Nahon et al. [10] enhanced the virtual reality headset experience with augmented virtuality using an off-the-shelf Kinect for Windows v2 to inject some reality in the virtuality. In 2019, Von et al. [20] use augmented virtuality technology for spotting passersby in room scale virtual reality. In 2021, Gonzalez et al. [6] proposed an advanced teleoperation and control system for industrial robots based on augmented virtuality and haptic feedback. At the same time, augmented virtuality attracts particular interests from researchers working on hazard-related safety trainings programs recently. In 2013, Chen et al. [5] proposes AVES, a safety training augmented virtuality environment for construction hazard recognition and severity identification. In 2014, Albert et al. [1] use AV to enhance construction hazard recognition with high-fidelity training scenarios. In 2018, Neges et al. [11] devised an AV interface for maintenance training simulation under various stress conditions. In this work, calibration from real object towards virtual object has been calculated accurately to achieve an immersive simulation in VR according to user's operations on real object. In 2020, Bhandari et al. [2] conduct a study using augmented virtuality to examine how emotions influence construction-hazard identification, risk assessment, and safety decisions.

Music Performance in eXtended Reality (XR). Recently, eXtended Reality (XR) technologies including VR, AR and MR have been widely adopted in virtual music performance, music performance evaluations, and music performance anxiety training. In

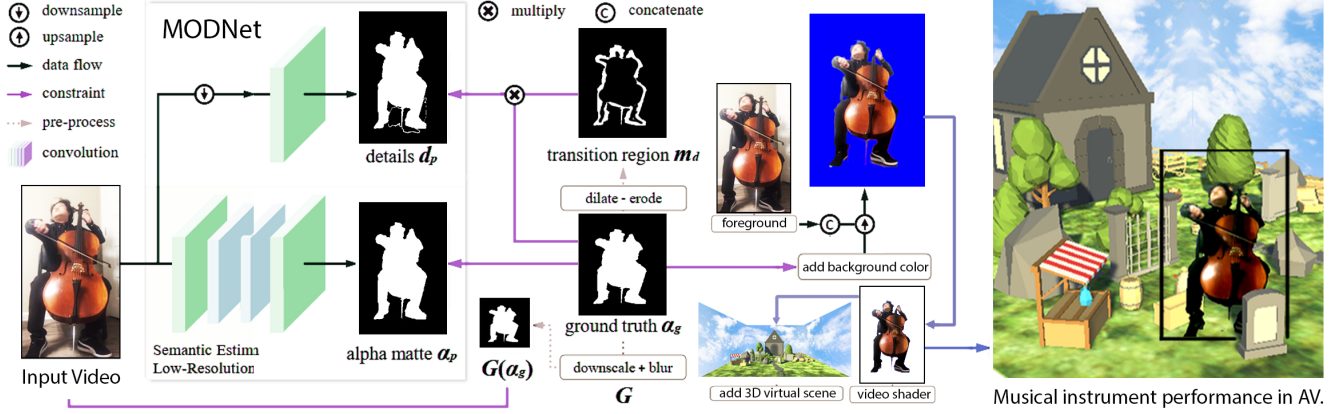


Figure 3: Overview of our technical approach.

2014, Keebler et al. [8] examined a novel augmented reality guitar learning system according to the current theories of embodied music cognition. In 2016, Bissonnette et al. [3] studied the musical instrument players' performance anxiety and quality of performance during virtual reality exposure training. From the feasibility study conducted by Orman et al. [12] in 2017, it showed that students using the immersive Virtual Reality Learning Environments (VRLE) with head tracking demonstrated greater conducting skill improvement than those not using virtual reality. In 2018, Trujano et al. [17] proposed ARPiano: an efficient piano learning interface using augmented reality. Later, networked music performance in virtual reality has been systematically studied by Loveridge et al. [9] in 2020. The same year, Petrovic et al. [14] explored how do the augmented and virtual reality web applications contribute to the music stage performance. Van et al. [19] studied the effects of virtual reality on music performance anxiety among university-level music majors. Most recently, Van et al. [18] proposed a methodological framework for assessing social presence in music interactions via virtual reality in 2021. All these research works mentioned above show positive feedback that there is potentially a great market in applying eXtended Reality (XR) technologies onto music performance. Unfortunately, no existing research has given it a shot to use augmented virtuality in musical instrument performance so far. Therefore, we open a novel topic in applying AV to the music art performance industry.

3 TECHNICAL APPROACH

Overview. Figure 3 shows the overview of our technical approach for synthesizing musical instrument performance in augmented virtuality. First, given the musician's prerecorded performance anywhere with any camera device (can be low resolution), video is sent to the MODNet for video matting. After the video matting process, the output video is written by synthesizing the foreground from the video and the background with a color that has not appeared in the video (in this case, it is blue). The last step is sending the colored video matting to the 3D graphics renderer. Then, given the video GPU shader defined by us, the input video can be rendered realistically in the 3D virtual scene in real-time. So that the animation of the musician's prerecorded performance is transformed into the performance in augmented virtuality.

Video Matting. In order to remove the background from the musician's input video, we applied semantics image segmentation on each frame in the input video using a deep learning approach. By training a deep convolutional neural network that can classify which pixel on the image belongs to the foreground and which belongs to the background, output signals values between 0 and 1 can be synthesized for each pixel and form a new gray image called alpha matte. In this figure, the white shape represents the foreground while the black shape represents the background that is segmented from the input image. In our approach, we use a state-of-the-art model known as MODNet. Unlike traditional image segmentation methods which generate binary images as the segmentation results, MODNet generates image mattings that are referring to the opacity of the foreground. Figure 3 shows an abbreviated version of the architecture of MODNet. For more details of MODNet, please refer to the article written by Ke et al. [7]. Unlike other image matting algorithms, MODNet considers two major steps: (1) Semantic estimation that outputs a coarse foreground mask. (2) Detail prediction that outputs a fine foreground mask. The last step is the semantic-detail fusion that blends the features from both the first two layers. According to the experimental results, MODNet works well on both human matting and music instrument matting for most of the cases.

Video Shading. In order to present the musician in the virtual environment naturally, we consider two major factors in the video shader: (1) alpha map smoothing and (2) realistic rendering. For alpha map smoothing, we smooth the edge of the video matting result through a smoothstep function $S(x, s_0, s_1)$ implemented in the GL/SL shader language using a cubic Hermite interpolation after clamping so that $S(x, s_0, s_1) = 3[(x - s_0)/(s_1 - s_0)]^2 - 2[(x - s_0)/(s_1 - s_0)]^3, x \in [s_0, s_1]$; Otherwise, $S(x, s_0, s_1) = 0$. Then alpha map $\alpha(u, v)$ measured by difference between the pixel color of input video $C_{in}(u, v)$ and the background color C_{bg} is: $\alpha(u, v) = S(\|C_{in}(u, v) - C_{bg}\|, s_0, s_1)$, where $(u, v) \in [0, 1]^2$ are the texture coordinates, $s_0 = 0.3$ and $s_1 = 0.35$ are two threshold empirically set for smooth visual effects. For achieving a more realistic visual effect that the virtual environment in the background matches well with the musician in the foreground, we consider the realistic rendering lighting equation. Due to the missing information of the normal direction in each pixel in the video, we approximate the



Figure 4: Experimental results of musical instrument performance in Augmented Reality (AR).

lighting equation using the ambient lighting model: $C_{out}(u, v) = \alpha(u, v)[(1 + \beta)(C_{in}(u, v) + C_{am}) - C_{sh}]$ where C_{am} is the approximate ambient light color, β is the ambient light intensity, and C_{sh} is the shadow color which is typically light gray.

4 EXPERIMENTAL RESULTS

Different Scenes. Figure 4 shows the experimental results of musical instrument performance in Augmented Reality (AR). In this case, a cello player is playing three different songs with different

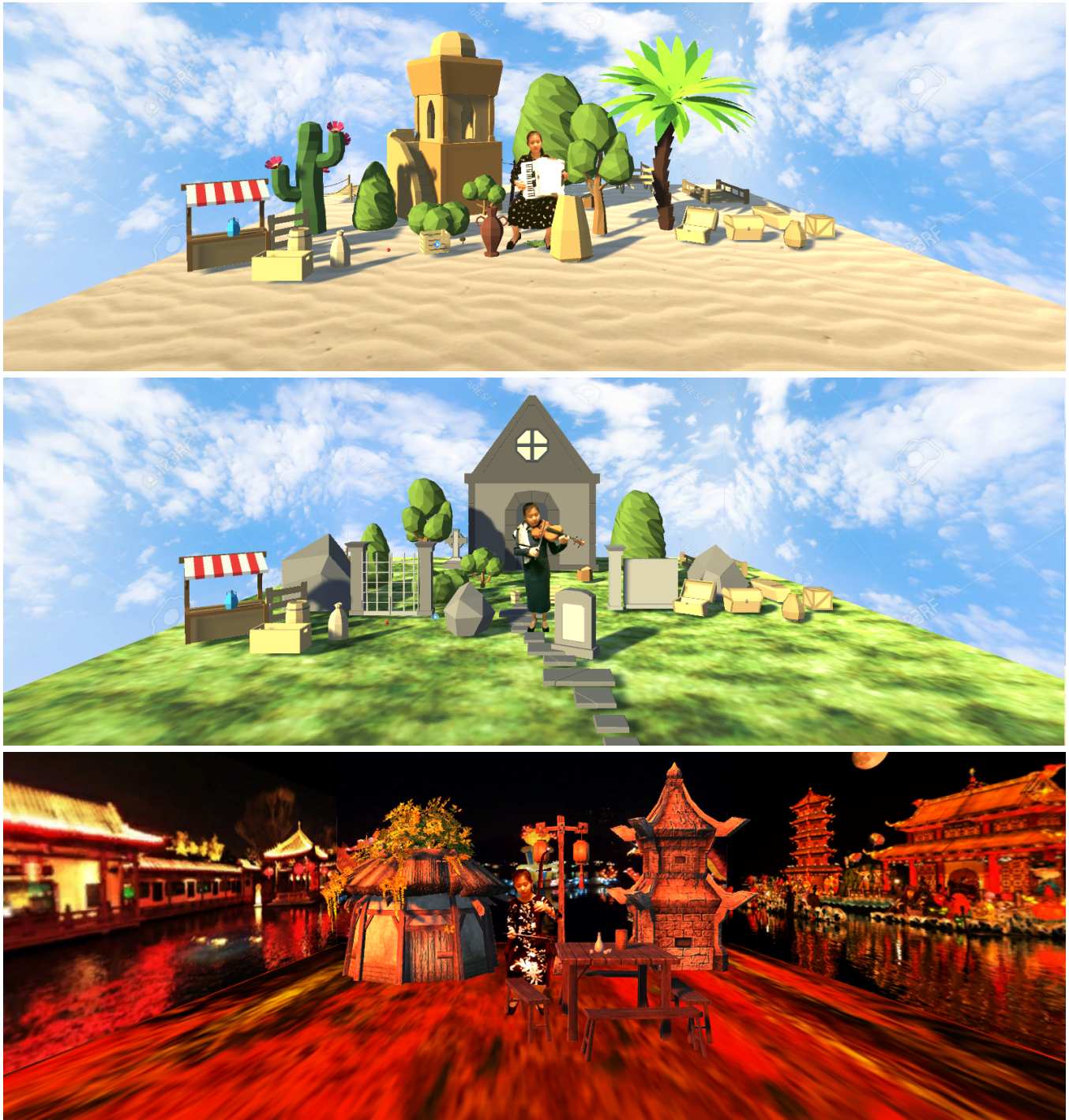


Figure 5: Performance in augmented virtuality with different types of musical instruments.

3D scenes as augmented virtual environments. The implementation of our proposed approach is based on Unity 3D with the 2019 version. We conducted the experiments with the hardware configurations containing Intel Core i5 CPU, 32GB DDR4 RAM, and NVIDIA GeForce GTX 1650 4GB GDDR6 Graphics Card. The player

of the video is implemented with a game object behavior class called *VideoPlayer* included in Unity 3D Editor by default. The video shader is implemented according to our approach with GL/SL Unlit and is attached along with a texture renderer to new material. That material is directly dragged onto a quad object in the 3D scene

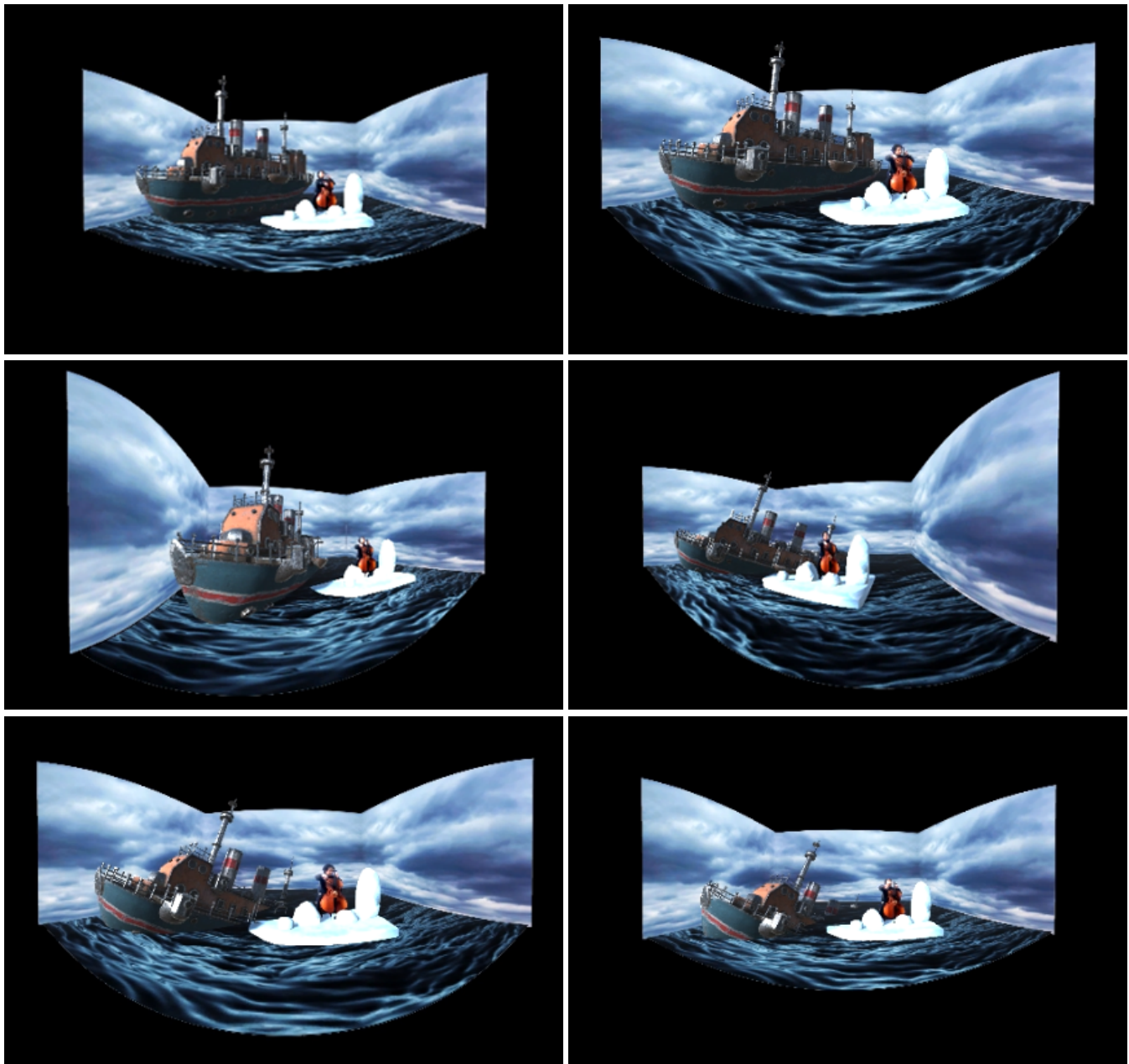


Figure 6: Performance in augmented virtuality is rendered with 360 VR video.

used to play the music performance. As we can see, the cello player looks natural in these virtual environments.

Different Instruments. Figure 5 shows another group of experimental results. In this case, a musician is playing three different songs with three different types of musical instruments within three different 3D scenes as augmented virtual environments. The musical instruments are accordion, violin, and erhu, respectively. This result shows the robustness of our approach to matting different types of instruments. As we can see, both accordion, violin, and erhu are clearly segmented from the background with very high quality.

With these three different styles of instruments, we designed three different types of 3D virtual scenes, they are: a Muslim-style village is corresponding to an accordion that is playing a Xinjiang Uygur song, a western catholic-style park is corresponding to a violin that is playing an England song, and a Chinese festival-style night scene is corresponding to an erhu playing a traditional Chinese song.

360 VR Video. We test our interface with 360 VR video settings. As shown in Figure 6, a cello player is playing a song on a virtual ice island where there is a ship nearby. This music is a piece extracted from the *Titanic Theme*. The virtual scene is set up to mimic the

story of the movie *Titanic* where the ship is sinking while the ice island is floating on the sea. The animations of the floating island and ships are simulated with sine functions. The sinking effect is simulated through a translation function (sinking) and a rotation function (tilting). The camera's moving trajectory is manually set up with the timeline control embedded in the Unity Editor's Recorder Package. Then according to the script, the cubemaps from both the left eye and the right eye are rendered on a texture renderer called Equirect. After the animation sequence is rendered in the play mode, the image sequence of the 360 VR video is generated. After a quick operation in the Adobe Premiere Pro, the 360 VR video is rendered. Those subfigures in Figure 6 show six different frames within the 360 video in chronological order. For each subfigure, the cubemap rendered from the camera on the left eye is presented. Due to the similarity between the cubemaps from the left eye and the right eye, the right eye's cubemaps are not presented in this paper. Video demo of these experimental results can be accessed with this link: <https://youtu.be/IHGKqYRZRVI>.

5 CONCLUSION

In this paper, we present the preliminary study results about musical instrument performance in Augmented Virtuality (AV). We design a full implementation pipeline to convert an arbitrary input video of a musician's performance to a well-designed 3D performance in augmented virtuality. During the implementation, we have considered several important computer vision and computer graphics techniques such as deep learning-based real-time video matting techniques using MODNet and realistic rendering techniques for video shading. During the experiments, we have invited several musicians to play different songs within different virtual environments or playing different songs with different musical instruments. During the experiments, we augment the players' performance with different virtual scenes according to the corresponding styles of the songs. All results show acceptable visual effects, attractive natures of augmented virtuality, and great potential for applying augmented virtuality for musical instrument performance in daily life.

Given the convenience of implementing and applying our proposed technical approach to daily use, there is great potential to extend our proposed approach to popular social media where musicians are more than welcomed such as Tiktok, Vlog, YouTube, etc. Its commercial potential can be very high if a group of professional people in a 3D design team can be organized to design and sell 3D virtual theaters for AV music performances. Besides, our proposed approach can be extended to more interesting art performances other than musical instrument performances such as dancing performances, talk shows, and singing shows. As for the technical improvement, in future work, we are aiming at improving the video matting approach in real-time on the VR headset such as Oculus quest, where the video input is no more pre-recorded musical instrument, rather, it is achieved while the user is looking around with VR headset put on. This can achieve an even more immersive experience for the audience to watch the music instrument performance in AV. But unfortunately, so far there is no such technology, and needs lots of effort in exploring this direction. Besides, a large-scale user study is waiting for us to analyze the comparison between our proposed interface for music performance in AV and

other interfaces such as VR and AR music performance. In that way, we may claim with stronger confidence that our AV interface is more immersive for musical instruments performance than other interfaces such as VR and AR.

REFERENCES

- [1] Alex Albert, Matthew R Hallowell, Brian Kleiner, Ao Chen, and Mani Golparvar-Fard. 2014. Enhancing construction hazard recognition with high-fidelity augmented virtuality. *Journal of Construction Engineering and Management* 140, 7 (2014), 04014024.
- [2] Siddharth Bhandari, Matthew R Hallowell, Leaf Van Boven, Keith M Welker, Mani Golparvar-Fard, and June Gruber. 2020. Using augmented virtuality to examine how emotions influence construction-hazard identification, risk assessment, and safety decisions. *Journal of construction engineering and management* 146, 2 (2020), 04019102.
- [3] Josiane Bissonnette, Francis Dubé, Martin D Provencher, and Maria T Moreno Sala. 2016. Evolution of music performance anxiety and quality of performance during virtual reality exposure training. *Virtual Reality* 20, 1 (2016), 71–81.
- [4] Maria Luz Caballero, Ting-Ray Chang, María Menéndez, and Valentina Occhialini. 2010. Behind: augmented virtuality gestural interaction for mobile phones. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 451–454.
- [5] Ao Chen, Mani Golparvar-Fard, and Brian Kleiner. 2013. SAVES: A safety training augmented virtuality environment for construction hazard recognition and severity identification. In *Proceedings of the 13th International Conference on Construction Applications of Virtual Reality*, London, UK. 373–84.
- [6] Claudia González, J Ernesto Solanes, Adolfo Muñoz, Luis Gracia, Vicent Gírbés-Juan, and Josep Tornero. 2021. Advanced teleoperation and control system for industrial robots based on augmented virtuality and haptic feedback. *Journal of Manufacturing Systems* 59 (2021), 283–298.
- [7] Zhanghan Ke, Kaican Li, Yurou Zhou, Qiuhua Wu, Xiangyu Mao, Qiong Yan, and Rynson WH Lau. 2020. Is a Green Screen Really Necessary for Real-Time Portrait Matting? *arXiv preprint arXiv:2011.11961* (2020).
- [8] Joseph R Keebler, Travis J Wiltshire, Dustin C Smith, Stephen M Fiore, and Jeffrey S Bedwell. 2014. Shifting the paradigm of music instruction: implications of embodiment stemming from an augmented reality guitar learning system. *Frontiers in psychology* 5 (2014), 471.
- [9] Ben Loveridge. 2020. Networked music performance in Virtual Reality: current perspectives. (2020).
- [10] David Nahon, Geoffrey Subileau, and Benjamin Capel. 2015. "Never Blind VR" enhancing the virtual reality headset experience with augmented virtuality. In *2015 IEEE Virtual Reality (VR)*. IEEE, 347–348.
- [11] Matthias Neges, Stefan Adwernat, and Michael Abramovici. 2018. Augmented virtuality for maintenance training simulation under various stress conditions. *Procedia Manufacturing* 19 (2018), 171–178.
- [12] Evelyn K Orman, Harry E Price, and Christine R Russell. 2017. Feasibility of using an augmented immersive virtual reality learning environment to enhance music conducting skills. *Journal of Music Teacher Education* 27, 1 (2017), 24–35.
- [13] Perrine Paul, Oliver Fleig, and Pierre Jannin. 2005. Augmented virtuality based on stereoscopic reconstruction in multimodal image-guided neurosurgery: Methods and performance evaluation. *IEEE transactions on medical imaging* 24, 11 (2005), 1500–1511.
- [14] Nenad Petrović. 2020. Augmented and Virtual Reality Web Applications for Music Stage Performance. In *2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*. IEEE, 33–36.
- [15] Holger Regenbrecht, Tim Lum, Petra Kohler, Claudia Ott, M Wagner, Wilhelm Wilke, and Erich Mueller. 2004. Using augmented virtuality for remote collaboration. *Presence* 13, 3 (2004), 338–354.
- [16] Kristian T Simsarian and Karl-Petter Akeson. 1997. Windows on the world: An example of augmented virtuality. (1997).
- [17] Fernando Trujano, Mina Khan, and Pattie Maes. 2018. ARPiano efficient music learning using augmented reality. In *International Conference on Innovative Technologies and Learning*. Springer, 3–17.
- [18] Bavo Van Kerrebroeck, Giusy Caruso, and Pieter-Jan Maes. 2021. A methodological framework for assessing social presence in music interactions in virtual reality. *Frontiers in Psychology* 12 (2021).
- [19] Marise van Zyl. 2020. The Effects of Virtual Reality on Music Performance Anxiety Among University-Level Music Majors. *Visions of Research in Music Education* 35 (2020).
- [20] Julius von Willich, Markus Funk, Florian Müller, Karola Marky, Jan Riemann, and Max Mühlhäuser. 2019. You invaded my tracking space! using augmented virtuality for spotting passersby in room-scale virtual reality. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 487–496.