Multimedia in Virtual Reality and Augmented Reality

Shu-Ching Chen, *Florida International University, Miami, FL, 33199, USA*

Multimedia is one of the key drivers improving virtual reality and augmented reality (VR/AR), which are promising to reform human–computer interaction in the future with lower-cost and all-in-one headsets containing powerful hardware. Advances in multimedia research on video compression and human–computer interfaces have further enhanced the immersion and efficiency of experiences on the platform. However, many VR/AR experiences are still very difficult to build using traditional engineering methods and many available behavioral and biometric data have not been well explored. Further research in the multimedia community is needed to enhance the usefulness of these systems, with potential in affective learning, resource generation, and developer tools.

Virtual reality and augmented reality (VR/AR) could be considered as one of the key technologies for the next generation of a human–computer interaction. VR/AR headset technology has become powerful enough that many traditional mobile or PC applications can now run on such headsets. Multimedia has played an important role to enable VR/AR technologies. To enable VR/AR devices to display the high-resolution virtual environments to the user seamlessly, immersive content needs to be effectively and efficiently projected and displayed on a virtual three-dimensional spherical surface. Continuous efforts from the multimedia community have led to the development of video-coding techniques, including layered video coding,[1] entropy equilibrium optimization,[2] etc. While the technology is ready for personal use and single-user applications, effective compression, and efficient transmission to enable interconnected VR/AR remains a fundamental challenge and further research is required.

While the users are able to see high-quality videos using VR/AR devices, their behaviors and surrounding environment are captured by the devices using cameras and sensors as well. So, multimedia tools and techniques can be applied to improve user experiences. Specifically, the multimedia data collected by the device can be used to facilitate user's interactions with the VR/AR environment. Advanced deep neural networks have been utilized to analyze the multimedia data collected by VR/AR devices to recognize patterns, such as speech,[3] hand postures, and gestures[4] to interact with the applications, removing the need for conventional controllers. However, the usability and reliability of these techniques to process data remain the key problem to allow for smooth and natural interaction in the VR/AR environment.[5]

VR/AR also provides access to user data that were previously difficult to collect, such as hand pose, head pose, eye-tracking, image, and audio data. These data types can provide great insights into a user's status, which could benefit domains, such as affective learning. Many VR/AR experiences mainly consist of rule-based systems utilizing the immersive nature of the platform, particularly in domains such as education. Such experiences have continuously been found to help improve user engagement,[6] which can be quite challenging using traditional online platforms, especially with hands-on operational work[7] or instructions to younger children such as those with disabilities.[8] However, others have shown that significant numbers of participants in certain VR/AR environments may actually learn less due to the distraction on the platform, even when they are engaged.[9] With the additional data that VR/AR hardware can allow researchers to collect passively, it may be possible to integrate these data into multimodal affective learning systems like those described by Verma *et al.*[10], Tao *et al.*[11] to modify experiences dynamically based on a user's

1070-986X © 2021 IEEE
Digital Object Identifier 10.1109/MMUL.2021.3086275
Date of current version 30 June 2021.

predicted needs and emotional state, keeping user engagement while augmenting learning.

The immersive nature of the platform implies that many assets and details are necessary to construct a realistic environment and a good sense of immersion, leading to enormous time and efforts developing assets in three-dimensional (3-D) space to accompany VR/AR experiences. This problem hinders developing new applications and immigrating existing ones for the VR/AR environment. Research in multimedia could also prove quite useful to mitigate and facilitate content generation for the VR/AR environments. Continued research into converting images into 3-D assets,[12] natural language descriptions into environments,[13] and the generation of audio[14] can all help minimize the work needed to create such environments for any use-case. To ensure that these generated environments are not simply static scenes, more research needs to be done for natural language-based code generation since current methods[15,16] lack those datasets needed to create robust models that could be utilized to generate functionality within the VR/AR environments.

Another challenging problem faced by VR/AR would be integrating the aforementioned multimedia methods and many other tools to allow developers to take advantage of the sensors and data available to them. Platforms such as Unity are extremely useful to create environments using traditional methods of asset creation, as well as providing some infrastructure to integrate the artificial intelligence models directly into projects.[17] However, it can still take a very long time to create VR/AR experiences. By creating frameworks and tools, it could be possible to allow users to generate fully functional environments themselves without the need for extensive computer science knowledge and avoiding lengthy development timelines.

In conclusion, VR/AR is a high-impact platform for the research and development of novel multimedia techniques and shows incredible promise in improving user outcomes in various domains. Advances in compression and interaction technologies as well as the multimedia data that can be collected have greatly improved the usability of the platform for user applications. However, more research needs to be done to take full advantage of the platform. VR/AR has proven to be a modality that can drive improved engagement within users, and further research into directions, such as affective learning, content generation, and development platforms can help fully realize the potential of the platform.

REFERENCES

1. A. T. Nasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using layered video coding," in *Proc. IEEE Virtual Reality*, 2017, pp. 347–348, doi: 10.1109/VR.2017.7892319.
2. Y. Zhou, L. Tian, C. Zhu, X. Jin, and Y. Sun, "Video coding optimization for virtual reality 360-degree source," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 118–129, Jan. 2020, doi: 10.1109/JSTSP.2019.2957952.
3. D. Hepperle, Y. Wei, A. Siess, and M. Wölfel, "2D, 3D or speech? A case study on which user interface is preferable for what kind of object interaction in immersive virtual reality," *Comput. Graph.*, vol. 82, pp. 321–331, 2019, doi: 10.1016/j.cag.2019.06.003.
4. K. M. Sagayam and D. J. Hemanth, "Hand posture and gesture recognition techniques for virtual reality applications: A survey," *Virtual Reality*, vol. 21, no. 2, pp. 91–107, 2017, doi: 10.1007/s10055-016-0301-0.
5. H. Tang, W. Wang, D. Xu, Y. Yan, and N. Sebe, "GestureGAN for hand gesture-to-gesture translation in the wild," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 774–782, doi: 10.1145/3240508.3240704.
6. B. I. Edwards, K. S. Bielawski, R. Prada, and A. D. Cheok, "Haptic virtual reality and immersive learning for enhanced organic chemistry instruction," *Virtual Reality*, vol. 23, no. 4, pp. 363–373, 2019, doi: 10.1007/s10055-018-0345-4.
7. P. Wang, P. Wu, J. Wang, H.-L. Chi, and X. Wang, "A critical review of the use of virtual reality in construction engineering education and training," *Int. J. Environ. Res. Public Health*, vol. 15, no. 6, 2018, Art. no. 1204, doi: 10.3390/ijerph15061204.
8. H. H. S. Ip *et al.*, "Enhance emotional and social adaptation skills for children with autism spectrum disorder: A virtual reality enabled approach," *Comput. Educ.*, vol. 117, pp. 1–15, 2018, doi: 10.1016/j.compedu.2017.09.010.
9. G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learn. Instruct.*, vol. 60, pp. 225–236, 2019, doi: 10.1016/j.learninstruc.2017.12.007.
10. M. Verma, S. K. Vipparthi, and G. Singh, "AffectiveNet: Affective-motion feature learning for micro expression recognition," *IEEE MultiMedia*, vol. 28, no. 1, pp. 17–27, Jan./Mar. 2020, doi: 10.1109/MMUL.2020.3021659.
11. Y. Tao *et al.*, "Confidence estimation using machine learning in immersive learning environments," in *Proc. IEEE Conf. Multimedia Inf. Process. Retrieval*, 2020, pp. 247–252, doi: 10.1109/MIPR49039.2020.00058.

12. B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 405–421, doi: 10.1007/978-3-030-58452-8_24.
13. Q. i. Chen, Q. i. Wu, R. Tang, Y. Wang, S. Wang, and M. Tan, "Intelligent home 3D: Automatic 3D-house design from linguistic descriptions only," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12625–12634, doi: 10.1109/cvpr42600.2020.01264.
14. S.Ö. Arık *et al.*, "Deep voice: Real-time neural text-to-speech," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 195–204.
15. B. Wei, G. Li, X. Xia, Z. Fu, and Z. Jin, "Code generation as a dual task of code summarization," in *Proc. Adv. Neural Inf. Process. Syst. Neural Inf. Process. Syst.*, 2019, pp. 1–11.
16. Z. Sun, Q. Zhu, Y. Xiong, Y. Sun, L. Mou, and Lu Zhang, "TreeGen: A tree-based transformer architecture for code generation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 5, 2020, pp. 8984–8991, doi: 10.1609/aaai.v34i05.6430.
17. A. Juliani *et al.*, "Unity: A general platform for intelligent agents," pp. 1–28, 2018, *arXiv:1809.02627*.

SHU-CHING CHEN is currently a Professor with Florida International University, Miami, FL, USA. Contact him at chens@cs.fiu.edu.