2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)

AVoidX: An Augmented VR Game

Rafail Athanasoulas * Prodromos Boutis † ‡ Anargyros Chatzitofis † § Alexandros Doumanoglou † ¶
Petros Drakoulis † ‖ Leonidas Saroglou † ** Vladimiros Sterzentsenko † †† Nikolaos Zioulis † ‡‡
Dimitrios Zarpalas §§ Petros Daras ¶¶

Center for Research and Technology Hellas, Information Technologies Institute, Visual Computing Lab

Figure 1: AVoidX in-game screenshots. The user aims to avoid the coming carved walls by striking appropriate poses. Collisions are driven by the user's volumetric representation instead of higher level metadata or projection-based techniques.

Abstract

The recent advances in real-time volumetric capturing enable new interaction paradigms in virtual environments. Volumetric representations can elevate the feeling of presence, and therefore increase the user immersion by emplacing them within virtual environments allowing them to occupy space. This research demonstration aims to showcase the possibilities offered by real-time volumetric capturing in an Augmented VR context. More specifically, a novel single-player obstacle avoidance game is presented where the users try to avoid collision with incoming carved walls by moving their actual bodies.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality; Computing methodologies—Computer graphics—Graphics systems and interfaces—Mixed / augmented reality; Computing methodologies—Artificial intelligence—Computer vision—Camera calibration; Computing methodologies—Artificial intelligence—Computer vision—Reconstruction; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 Introduction

The advent of the original Microsoft Kinect in 2010 promised to revolutionize the way humans interact with digital content through its depth sensing and skeleton tracking capabilities. However, partial noisy depth maps and occlusion-sensitive motion tracking prevented adoption in more mature solutions. Nevertheless, depth sensors are reaching a technology readiness level that allows for their embedding in advanced systems.

In this work, we aim to show the potential of integrated color and depth sensors that allow for partial (single-view) volumetric data capture. By leveraging multiple sensors surrounding the user we can volumetrically reconstruct them and acquire digitized human representations. We develop AVoidX, an Augmented VR [4] game that emplaces volumetrically captured users in a game environment that seeks to accentuate the feeling of presence. The goal of this demonstration is to showcase the combined effect of **i) visual presence**, enabling a realistic representation of the users using their own actual appearance in the VR environment; **ii) spatial presence**, that accurately emplaces users within the virtual world in a life size scale; and **iii) interactive presence**, enabling the bridging of the virtual world with the physical space that the user occupies.

2 System Overview

AVoidX requires a technologically multi-disciplinary approach as the overall system comprises an end-to-end volumetric video production pipeline. Multiple sensors capture color and depth data, which are fused in a live manner into a suitable representation. This representation gets compressed, transmitted and eventually embedded and rendered within a virtual game environment. The following subsections describe the respective sub-systems that comprise the main architectural blocks of the aforementioned pipeline, using a high-level perspective and composing the technical outline of AVoidX, presented in Fig. 2.

2.1 Portable Volumetric Capture

A publicly available[1], decentralized multi-view sensor acquisition sub-system [8] is used that consists of two main modules: the *remote eyes* which serves data from a single sensor (*i.e.* viewpoint) and the *multi-view client* that consolidates them and is responsible for their spatial and temporal orchestration. The latter functionalities rely on a structure-based multi-sensor calibration method [7] and IEEE PTP 1588 protocol [3] (as the acquisition sub-system runs on a local network), respectively. The multi-view client is responsible for publishing temporally aligned frames as well as the spatial alignment metadata for each of the served sensor streams. Information (*i.e.* control and data) exchange between the remote eyes and the multi-view client is accomplished through a pub-sub broker.

*e-mail: rafail.athan@gmail.com – Game Development
†Equal contribution (in alphabetical order) – System Development
‡e-mail: prod@iti.gr
§e-mail: tofis@iti.gr
¶e-mail: aldoum@iti.gr
‖e-mail: petros.drakoulis@iti.gr
**e-mail: saroglou@iti.gr
††e-mail: vladster@iti.gr
‡‡e-mail: nzioulis@iti.gr
§§e-mail: zarpalas@iti.gr
¶¶e-mail: daras@iti.gr

[1] `https://github.com/VCL3D/VolumetricCapture`

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.00271

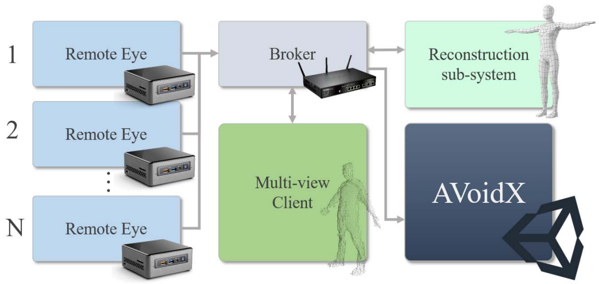

Figure 2: High-level diagram of the end-to-end volumetric video production pipeline and the AVoidX game.

From a hardware point of view, four sensor data serving agents are used. Each one serves live streams from an Intel RealSense D415 [5] sensor, each connected to a mini-PC host that is mounted on the same tripod the sensors are. These are placed at 90^o angle intervals perimetrically covering a capturing space where the user will be placed. A laptop hosts the multi-view client, while connectivity is wired via a 1Gbps switch. In addition, custom PCB board based synchronization cables ensure synced hardware triggering for frame acquisition from all the sensors, while commodity packaging boxes are used for assembling the calibration structure. Overall, the capturing system is low-cost and portable, an enabling factor for such a demonstration.

2.2 Live User Representation

The spatially and temporally aligned multi-view data are then received from the broker and fused by the reconstruction sub-system. A fully parallel Fourier-based 3D reconstruction method is employed [1] that allows for real-time processing on reasonable voxel grid resolutions even on modest gaming laptops. This voxel based volumetric representation is meshed into a geometric representation via Marching Cubes [6]. It is also accompanied by per-vertex texture blending metadata, estimated in tandem, which are then compressed in an intra-frame manner, given the real-time restriction. These, in addition to the acquired textures, are streamed to the system's broker and offer a live 3D user representation that can faithfully reconstruct the appearance and shape of the user inside a virtual environment. Multi-view texture blending allows us to mask the geometric quality defects (*i.e.*low resolution grid and lack of details), and offer a realistic digital representation of the users in real-time.

3 Real-time Demonstration

In this demonstration, we showcase the aforementioned volumetric capture system and the newly introduced AVoidX game, which is open-source and publicly available[1]. The full setup for the demonstration is illustrated in Fig. 3. Users are reconstructed in real-time and play the game using the volume of their own bodies. We use a $64 \times 128 \times 64$ voxel grid resolution to achieve real-time rates, even with constrained processing (*i.e.* laptop). The game's goal is to avoid incoming walls, which are carved with empty regions, indicating the poses that the users need to strike in order to safely avoid them. With respect to the game logic, the users earn points by successfully avoiding the walls and gathering virtual coins, while the top players of the game are shown on a leader board. In this context, the user's shape and motions within the capturing area are directly mapped into the virtual environment. Interestingly, the same applies in the backward direction, as the incoming obstacles that the users need to avoid virtually, occupy space in the physical world since the users

[1]`https://github.com/VCL3D/AVoidX`

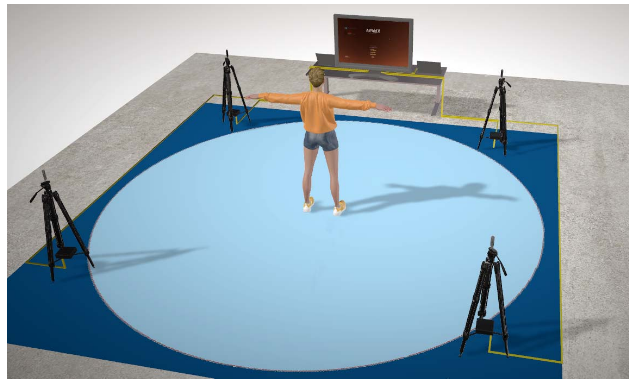

Figure 3: Four sensors perimetrically cover the capturing space where the user is placed watching and playing the game on a large screen.

seek to avoid them, and thus, move accordingly. This makes for an interesting bridging of the virtual and real worlds that will elevate the feeling of presence for the users within the game and similarly impose the virtual world's directives in the actions performed in the real world. This is a novel concept that goes beyond previous Augmented VR experiences [2] that restrained the users' motions and limited their interactions in triggered based events and body posture based navigation. The AVoidX demonstration and in-game video can be found on `https://youtu.be/jt9EFqwH3Zg`.

Acknowledgments

We acknowledge financial support by the H2020 EC project Hyper360 (GA 761934).

References

[1] D. S. Alexiadis, A. Chatzitofis, N. Zioulis, O. Zoidi, G. Louizis, D. Zarpalas, and P. Daras. An integrated platform for live 3d human reconstruction and motion capturing. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(4):798–813, 2016.

[2] K. Christaki, K. C. Apostolakis, A. Doumanoglou, N. Zioulis, D. Zarpalas, and P. Daras. Space wars: An augmentedvr game. In *International Conference on Multimedia Modeling*, pp. 566–570. Springer, 2019.

[3] J. Eidson and K. Lee. Ieee 1588 standard for a precision clock synchronization protocol for networked measurement and control systems. In *Sensors for Industry Conference, 2002. 2nd ISA/IEEE*, pp. 98–105. Ieee, 2002.

[4] A. Karakottas, A. Papachristou, A. Doumanoqlou, N. Zioulis, D. Zarpalas, and P. Daras. Augmented vr. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–1. IEEE, 2018.

[5] L. Keselman, J. Iselin Woodfill, A. Grunnet-Jepsen, and A. Bhowmik. Intel realsense stereoscopic depth cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–10, 2017.

[6] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *ACM siggraph computer graphics*, vol. 21, pp. 163–169. ACM, 1987.

[7] A. Papachristou, N. Zioulis, D. Zarpalas, and P. Daras. Markerless structure-based multi-sensor calibration for free viewpoint video capture. 2018.

[8] V. Sterzentsenko, A. Karakottas, A. Papachristou, N. Zioulis, A. Doumanoglou, D. Zarpalas, and P. Daras. A low-cost, flexible and portable volumetric capturing system. In *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 200–207. IEEE, 2018.