

A Framework for Cypher-Physical Human-robot Collaborative Immersive MR Interaction – Beaux Arts Ball 4.0

Risu Na^(⊠) and Haocheng Dai

University of California Los Angeles, Los Angeles, CA 90095, USA narisu960209@ucla.edu

Abstract. In this paper, we presented a human-robot collaborative mixed reality application – Beaux Arts Ball 4.0, in which a real-time interactive hybrid and physical architectural environment were designed and experienced through the tools and techniques of mixed reality, cypher-physical, teleoperation, telepresence, and automation. The application engaged the user and observer in a continuous loop of architectural transformation during the experience, where every type of sensory was blurred between physical and digital perception.

Keywords: Mixed reality · Avatars · Robotics · Human-robot collaboration

1 Introduction

In reference to the new virtual environments of industry 4.0, we proposed a contemporary interpretation of Beaux Arts Ball, a historic social environment where participants not only included people but also iconic buildings in the forms of costumes and other kinds of stage design elements. Beaux Arts Ball 4.0 is an interactive mixed reality environment, in the tradition of the Architecture League of New York's Ball, within the context of the 4th industrial revolution.

We explored how to bridge between the physical and digital world with a system of sensors in a spatial context, expanding upon the current forms of mixed reality experience. In this process, both the human body and robots were designed as an 'aggregated' character, whose behavior and performances helped build the 'aggregated' environment. The process culminated in an architecturally augmented robotic performance. An observer's position and point of view will be tracked in real-time to reveal the augmented environment, complete with avatars of the tele-present participant and digitally augmented physical robots. The digital avatar and augmented KUKA robots are actors in the synthetic scene, they interact with each other based on participants' input and distinct behavioral patterns through machine learning (Fig. 1).

The research was based on HTC VIVE VR system, a consumer-grade VR device, and two KUKA 6-axis KR150 articulating robotic arms. The main used platforms and programs were Unity 3D, Steam VR, and Autodesk MAYA in addition to Robot drivers that use UDP/IP User Datagram Protocol for exchanging data with robots.



Fig. 1. Telepresence composition diagram

2 Related Works

Originating back to the Ecole des Beaux Arts Ball in France, the ball included costume design, cross-dressing acts, often from human to building, iconic floats on river Seine and other antics that question and re-interpreted the presence of human and require the participation of design forms in a social space (Fig. 2). The design elements in the forms of costumes and sets were intended to be the participants, contributing to the complexity of social interactions through the creation of an alternative reality that was visually and experientially different in its interpretation of the human body and its adjacency to other bodies. The spirit of Beaux Arts Ball and other historic costume parties where the human bodies were altered and merged with some conceptual design contents serves as a precedent to many social VR platforms today. These platforms constitute humanoid avatars for tele-present human participants in a virtual environment and automated non-physical elements in the form of simulation and interface to help streamline the whole communication process.

Virtual reality was used by architects for design concept presentation. In their paper, Schnabel, Wang, and Kvan stated that the virtual environment gave architects an opportunity to express and explore their imagination more easily [1]. Beyond adding a virtual entity to the real world view, VR technology also enhanced the collaboration between design team members [2]. The collaborative interface allowed the users to share the virtual space with the other party promoting collaboration. Some of the collaborative interfaces even used different viewports in VR and AR to support different collaborative roles [3]. However, many researches in VR and AR space just focused on collaboration between users in either only VR or AR environments. The most recent work that combined the VR and AR space for collaborative experience was CoVAR [14]. It used the depth data of the physical environments captured by AR devices to construct a VR environment for other users.



Fig. 2. Photo from 1931's Beaux arts architect ball [15]

The most common multi-user virtual environment approaches focused on representation and interaction purely within the virtual space. With the launch of low-cost head-mounted displays (HMDs), networked mixed reality environment have increased its popularity in the remote collaboration field. MR occupied the ranges of the continuum between pure real environments and purely synthetic virtual environments, by merging them together. Strauss et al. presented "Murmuring Fields" as a mixed reality shared environment installation for several users based on a decentralized network. In this scene, "Murmuring Field" is a sound reacting to the users' body movement, movement triggers sound in the virtual space which could be heard in physical space [4]. Georges and Cédric introduced a setup and framework of avatar staging theatrical mixed reality application. The research focused on the relationship between performer, avatar, and audiences in an environment mixing 3D digital space and physical 'traditional' staging space [5]. However, there were very limited researches focused on the mixed reality collaborative experience, where users/agents working on the same 'project' at the same time.

Human-robot collaboration was a new trend in industrial robotics as part of strategy industry 4.0. The main purpose of this strategy was to create an environment between humans and robots to work collaboratively. Even the technology was still in its infancy, some researchers developed applications regarding human-robot collaboration. Exquisite Corpus was an installation by Songwen Chung, in which she painted with her three robotic collaborators. In this performance, she explored a process of human & robotic co-creation [6]. In a series of papers, Dragone and Holz raised the idea that displaying a humanoid avatar upon the robot platform could help people understand the robots' status more effectively [7]. Similarly, Dragone and Holz presented a novel methodology, which combines the physical robot body and a virtual character displayed through a mixed reality overlay [8]. Our experiments complemented that work, further exploring the mixed reality robot/avatar design space.

In this paper, we established a human-robot collaborative mixed reality interaction to let users experience the real-time construction process of an 'aggregated environment'.

3 Methodology and the Main Procedure

Each participant wore HTC VIVE headset, which determined the exact x, y, z position and rotation of the participant's field of view within the ecosystem of sensors and streamed a camera in the virtual world to screen projections. The screens and projections represented a portal from the physical world to the digital world. Additionally, the robots were augmented as avatars, too. They reacted in real-time in the scene and work as separate entities.

3.1 Virtual Aggregated Environment

In nature, large masses of granular substances, like sands, stones, gravels, form the different kinds of landscape and objects through the process of erosion and accretion. We imitated this natural behaviour in architecture and created its own configuration [9].

In this application, we explored the construction of an aggregated spatial enclosures through designed granular materials, which consisted of a large number of particles. Three types of particles with different behaviours were defined, convex spheres, which could flow, cubicles, which could be used to form the edges, and convex hexapods, which could interlock to be self-supporting structures. These particles were not bonded together in this process, they only interacted through frictional contact. Such unbound granular structures revealed the unique property to be both stable as a solid material and reconfigurable like fluid.



wall + column + landscape

Fig. 3. Avatar behaviors and control diagram

The construction event was executed with the participation of an avatar and two augmented robots. During the experience, users could control avatar and robots to generate/shoot particles into pre-made transparent boundary container molds. We selected a few typical architectural elements, like columns, hexagon wall structures, and landscape, to visualize the whole process (Fig. 3).

3.2 Avatar

The avatar was a remote, spatial, and abstract representation of the user/participant in this event. It trans-located the physical users from the actual location into the scene. The

avatars were designed according to the formal and behavioral characteristics related to the 'aggregated' environment. During the 'construction scene' experience, users could use physical gestures like shaking-off, jumping, hitting to interact with the particles, or could use pre-programmed buttons to execute a specific task, such as generating particles on own body, shooting particles into the scene, etc. (Fig. 4).



Avatar generating particles

Avatar shaking off particles

Avatar shooting particles

Fig. 4. Avatar behaviors and control diagram

The system transformed the sensor data obtained from the user's head-mounted display (HMD) and two hand controllers in the lighthouse tracking system of HTC VIVE, which detected the exact location of the devices in a room-scale environment. All the sensor data were translated into position and orientation information in world coordination with commercially available Unity3D plug-in Open VR. The position and coordination of the user's head and two hands were updated in every frame to determine the velocity, direction, and other information for motion reconstruction.

A process was presented within which an individual avatar was created using a simple 3-point tracking system and moved with inverse kinematics. Considering in our application, all the interactions were implemented by using two hand controllers, users would pay more attention to the motions of upper limbs and communication latency rather than a full-body reconstruction accuracy. So, we selected Limb (IK algorithm), which came with the commercially available Unity 3D plug-in Final IK. It had a shorter communication time, better Mean Per Joint Position Error (MPJPE), and better Mean Per Joint Rotation Error (MPJRE) per Table 1 [10].

Method	MPJPE	MPJRE	Time
Limb	29.5 mm	67.9°	0.1 ms
CCD	54.7 mm	105.8°	0.8 ms
FABRIK	43.7 mm	88.4°	0.2 ms

Table 1. Comparison between different IK algorithms

During the experiment, we instructed the user to take a standard T-pose to take the world position of the left hand, right hand, head, and the height of virtual body was

assumed. We performed a random dancing experiment to verify the validity and effectiveness of our method. As shown in Fig. 5, some random poses in a free dancing process were shown to prove that our method could estimate feasible and stable behaviors. For the upper body, the IK worked well to reconstruct the arms and upper body motions according to the 3 points tracking data. For the lower body, the legs and hips are centered at head and chest, which helps to get a stable and natural lower body orientation.

The goal of this method was to reconstruct a visually appealing motion in real-time by 3 tracking points, the precision of the result is not our strength, since we don't have tracking data of the lower body. In the application, the user's view was blocked by HMD. They didn't see the difference between an avatar and themselves as long as the reconstruction motion is natural and stable. So it was worth minimizing the tracking device instead of improving the precision.



Fig. 5. Motion reconstruction capabilities, reconstruction results of random dancing

3.3 Robotic Movement

Two augmented robots were the remote and abstract representation of physical robots in this event. The augmented robots were also designed to work in connection with the aggregated environment, as they were set to be assistants of avatars in 'construction scene'. During the experience, on one hand, users could use the controller button to let one robot generate/shoot particles into the scene to help construction. On the other hand, the other robot was set to do one specific job programmed in the application, such as sweeping the particles already generated from robot and avatar to arrange the scene (Fig. 6).

A complete 3d model of KUKA robots was assembled and imported into Unity as a.fbx file extension. The model was imported as a tree of connected joints with constraints between, so that each part was a child of the previous part starting from the base point. After imported into Unity, each part of the robots including aggregated elements were assigned to different game objects, which was sorted into certain hierarchical order corresponds to the DOF links of real-world robots (Fig. 7)



Sweeping particles

Sweeping particles

Shooting particles

Fig. 6. Robot behaviors and control diagram



Fig. 7. Virtual 'Robot Cage' arrangement

There were two game object groups set for each robot, 'Robot_GRP' (actual game object seen in the scene) and 'Robot_Ghost' (invisible in the scene). Each group contained the same hierarchic tree structures of joints and game objects. In our case, the tracking data was extracted from left-hand controller of HTC Vive once per frame in Unity. These data were used to generate the robot's parameter which defined the actual robot's movement. So, we used a data filtering process to eliminate the outliers, which helped the robot move smoother and avoid collision because of unstable controller movement. And this process took place in 'Robot_Ghost'. The 'clean' tracking data was sent to 'Robot_GRP' to calculate joints rotation data, which were sent to the actual robot by UDP connection. In this case, one more layer of movement protection was added between actual robot and hand controller, and the movement of actual robot and virtual robot avatar matched perfectly. The formula of data filter was given in this equation:

$$R_Ai = a * R_Ai + b * R_A(i - 1)$$
 (1)

where:

 R_A_i : is the current position of target object. a: float effector. It helps to find a optimized path between position data.

By implementing IK (Inverse Kinematics) [11] based on the transformation matrix of the target object of 'Robot GRP', we computed a set of generalized parameters for each joint [12]. We attached the update scripts on every joint so that the position and orientation of every joint got updated in every frame. And the generalized joint parameters in the

virtual robot 'Robot GRP' were sent to the actual robot using an Ethernet connection (UDP/IP).

3.4 Portal

The projection and screen represented a portal to see from the physical world into the digital world, which provided a third-person view for this event to watch the performance of operator/avatar and robots.

As shown in Fig. 8, an HTC VIVE tracker was mounted above the projection screen, which defined the orientation of the screen into Unity 3D. And during the experience, users held another HTC VIVE tracker. Both trackers' orientation parameters defined the camera parameters in the virtual world, which helped users to peek into the virtual world through a portal.



Fig. 8. Portal – view from the physical world into the scene

4 Conclusion and Discussion

In this paper, we proposed an immersive mixed reality human-robot collaboration experience, by imitating the construction process of large-scale aggregated spatial enclosures. The application opened up a new possibility to an architectural immersive experience, where the sensory including the color, depth, materials, and geometries were constantly blurred between the physical and digital world.

Beaux Arts Ball 4.0 envisions and plans for future cyber-physical social environments where the participants are not limited to humans that are physically present in a particular space but constitute robots, artificially intelligent beings in the forms of sounds and simulations, digital and robotic avatars of other tele-present humans, and sensor-enhanced smart AI environments that are responsive and actively engaged with the social life of their context.

References

- 1. Schnabel, M.A., Wang, X., Kvan, T. Touching the untouchable: virtual-, augmented-, and reality. In: International Conference on the Association for Computer Aided Architectural Design Research in Asia (2008)
- Wang, X.: Mutually augmented virtual environments for architectural design and collaboration. In: Dong, A., Moere, A.V., Gero, J.S. (eds.) CAADFutures 2007, pp. 17–29. Springer, Dordrecht (2007). https://doi.org/10.1007/978-1-4020-6528-6_2
- Piumsomboon, T., Day, A., Ens, B., Lee, Y., Lee, G., Billinghurst, M.: Exploring enhancements for remote mixed reality collaboration. In: Mobile Graphics & Interactive Applications on SIGGRAPH Asia 2017 - SA 2017. ACM Press, New York (2017)
- 4. Strauss, W., et al.: Staging the space of mixed reality—reconsidering the concept of a multi user environment. In: Proceedings of the Fourth Symposium on Virtual Reality Modeling Language VRML 1999. ACM Press, New York (1999)
- 5. Gagneré, G., Plessiet, C.: Experiencing avatar direction in low cost theatrical mixed reality setup. In: Proceedings of the 5th International Conference on Movement and Computing. ACM, New York (2018)
- 6. Exquisite Corpus: Sougwen Chung (愫君). Sougwen.com (2019). https://sougwen.com/project/exquisite-corpus. Accessed 11 Apr 2021
- Dragone, M., Holz, T., O'Hare, G.M.P.: Mixing robotic realities. In: Proceedings of the 11th International Conference on Intelligent User Interfaces - IUI 2006. ACM Press, New York (2006)
- Dragone, M., Holz, T., O'Hare, G.M.P.: Using mixed reality agents as social interfaces for robots. In: RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. IEEE (2007)
- ICD Aggregate Pavilion: Uni-stuttgart.de. https://www.icd.uni-stuttgart.de/projects/icd-agg regate-pavilion-2018/. Accessed 11 Apr 2021
- Becher, A., Axenie, C., Grauschopf, T.: VIRTOOAIR: Virtual Reality TOOlbox for avatar intelligent reconstruction. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE (2018)
- Brandstötter, M.: Inverse kinematics of serial robots with wrist. MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/48468-inverse-kin ematics-of-serial-robots-with-wrist. Accessed 27 May 2021
- Murhij, Y., Serebrenny, V.: An application to simulate and control industrial robot in virtual reality environment integrated with IR stereo camera sensor. IFAC-PapersOnLine 52(25), 203–207 (2019)
- Dixken, M., et al.: Distributed, collaborative virtual reality application for product development with simple avatar calibration method. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE (2019)
- 14. Piumsomboon, T., Lee, Y., Lee, G., Billinghurst, M.: CoVAR: a collaborative virtual and augmented reality system for remote collaboration. In: ACM, p. 3 (2017)
- Famous Architects Dress as Their Famous New York City Buildings (1931). Open Culture (2014). https://www.openculture.com/2014/03/architects-dress-as-famous-buildingsthey-designed-in-vintage-1931-photo.html. Accessed 28 May 2021

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

