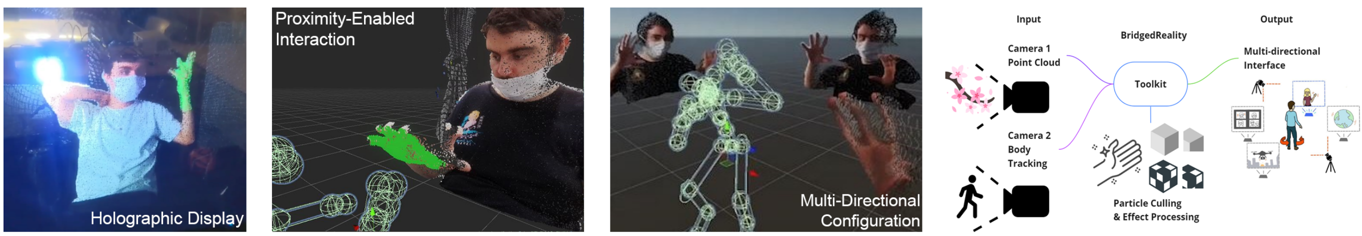
BridgedReality: A Toolkit Connecting Physical and Virtual Spaces through Live Holographic Point Cloud Interaction

Mark Armstrong
Lawrence Quest
Yun Suen Pai
mark@keio.jp
lawrence.quest@keio.jp
pai@km.keio.ac.jp
Graduate School of Media Design
Keio University
Yokohama, Japan

Kai Kunze
Kouta Minamizawa
kai@kmd.keio.ac.jp
kouta@kmd.keio.ac.jp
Graduate School of Media Design
Keio University
Yokohama, Japan



(a) (b) (c) (d)

Figure 1: (a) Tracked hand data, recolored and projected onto a large-scale holographic screen display (b) User's body is tracked in virtual space. Point cloud data within the probe's radius is modified in real time. (c) User's body is tracked in virtual space. Multiplicable RGB-D data surrounds the user from four angles, simulating the hybrid space. (d) System data input & output diagram.

ABSTRACT

The recent emergence of point cloud streaming technologies has spawned new ways to digitally perceive and manipulate live data of users and spaces. The graphical rendering limitations prevent state-of-the-art interaction techniques from achieving segmented bare-body user input to manipulate live point cloud data. We propose BridgedReality, a toolkit that enables users to produce localized virtual effects in live scenes, without the need for an HMD nor any wearable devices or virtual controllers. Our method uses body tracking and an illusory rendering technique to achieve large scale, depth-based, real time interaction with multiple light field projection display interfaces. This toolkit circumvented time-consuming 3D object classification, and packaged multiple proximity effects in a format understandable by middle schoolers. Our work can offer a foundation for multidirectional holographic interfaces, GPU simulated interactions, teleconferencing and gaming activities, as well as live cinematic quality exhibitions.

SA '21, December 14–17, 2021, Tokyo, Japan

ACM ISBN 978-1-4503-8687-6/21/12.
https://doi.org/10.1145/3476124.3488656

ACM Reference Format:
Mark Armstrong, Lawrence Quest, Yun Suen Pai, Kai Kunze, and Kouta Minamizawa. 2021. BridgedReality: A Toolkit Connecting Physical and Virtual Spaces through Live Holographic Point Cloud Interaction. In *SIGGRAPH Asia 2021 Posters (SA '21), December 14–17, 2021, Tokyo, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3476124.3488656

1 INTRODUCTION

Digital point cloud streaming, a relatively young visualization technique, enables the viewing of live remote content, including people, places, and various objects. Most implementations are egocentric, lacking multiple supported viewing angles [Zhang et al. 2020]. 3D object recognition from RGB-D input is notoriously time consuming [Xu et al. 2019], while interacting with RGB-D data in real-time is significantly more complex than working with static point clouds - new objects, and people, can freely enter and leave the scene at any time. Processing live data is also more computationally demanding than working with static point clouds. Complete data doesn't always exist between frames, and recent techniques for semantically disassembling a scene - typically for object detection - demonstrate varying levels of accuracy and low frame rates.

Graphically intensive softwares produce high-definition, artistic renderings of simple point cloud captures at cinematic frame rates. But to achieve interaction, the user (or their environment [Takenaka et al. 2021]) is usually equipped with trackers or virtual controllers. For some users, wearable devices and controllers are burdensome, which can be detrimental to the digital experience.

Past works suggest that user experience is also impacted by display scale [Yoshino et al. 2020]. Therefore, we propose *BridgedReality* which is a toolkit for bare-body point cloud interaction at large scale. The contributions of this work are the following:

(1) We introduce *BridgedReality*: A toolkit that averts the need for complex object detection algorithms, obviates the presence of virtual controllers, and can facilitate multi-directional RGB-D interaction through holographic spatial augmentation.
(2) We ran an in-the-wild preliminary exhibition to evaluate the our toolkit through user engagement and found high activity among young participants.
(3) We evaluate system impact and propose future applications.

2 METHOD

This work makes use of body tracking via the commercially available Microsoft Azure Kinect [1]. A camera captures live visual data from the physical world and mirrors it into virtual space. The toolkit then superimposes the spatial data of a tracked user (in a hybrid-space) into the virtual space where it can overlap the live RGB-D data. We showcase a variety of cross-reality proximity based interaction scenarios (Figure 1(a,b)).

Software Implementation. we use the Unity [2] game engine which features community made Azure Kinect Assets [3]. A VFX Graph generates a point cloud from camera input and renders it through the GPU. This is inherently sub-optimal for interactive use, since particles follow a pre-determined behavior, regardless of the scene content. The virtualization pipeline ends here, although many artistic works further enhance the quality of visualization through particle effects using node based programming [4]. This toolkit expands interactivity beyond the limitations of looped deterministic effects, by introducing a novel programming technique to allow users to control which section of the graph is impacted via spatial data.

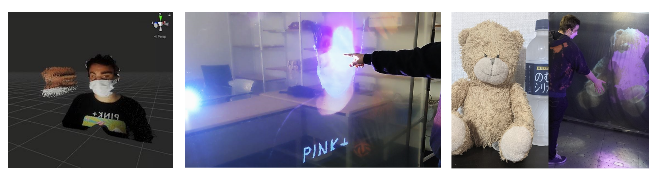
The Hidden Graph. Body tracking data from the same (or a secondary) camera, provides the system with multiple probes (e.g. the hand, head, foot) that supply spatial data. The challenge is to simulate "collision" between a selected joint and nearby point data. Collision detection in the CPU is slow with dense scenes. Collision detection post-GPU is impossible, as spatial location of point cloud data is irretrievable once sent to the GPU. Our technique is to visually simulate collision using a second, "hidden" VFX graph, that is mounted on any probe. This "hidden graph" performs particle culling across the entire capture with the exception of those particles within a defined radius of the probe (Figure 1(d)). Multiple probes can be attached to a single skeleton, allowing users to simultaneously perform distinguished bare-body interactions, even with data that is not of themselves (Figure 2(b). This approach preserves the bandwidth of data, and performance is not sacrificed for the sake of feature detection [Hosseini and Timmerer 2018].

[1]https://azure.microsoft.com/ja-jp/services/kinect-dk/
[2]https://unity.com/
[3]https://rfilkov.com/2019/07/24/azure-kinect-examples-for-unity/
[4]https://github.com/keijiro/Akvfx

Large Scale Displays. We use a large scale holographic screen [5] as our interface, which projects the processed RGB-D data. Users are then able to approach, and manipulate the data they see and make calibrations to their posture based on perceived information. This process generates a closed feedback loop where elements of physical space influence a virtual space, which then in turn influences the physical space through augmentation, creating a hybrid-space. In this space, the user can freely roam around and interact with one, or even multiple displays (Figure 1(c)) with no physical encumbrances, shifting the computational workload onto the environment. To demonstrate this concept, we built multiple holographic displays to show identical and distinct content, giving users the freedom to interact with their preferred interface or effect (Figure 2(a)).

3 PRELIMINARY EXHIBITION

We hypothesize that bare-hand interaction with varying virtual objects will strongly impact the perceived quality of interaction and immersion based on recent work [Tang et al. 2020]. To test this hypothesis we designed an in-the-wild exhibition to measure the engagement and interest of users to validate the effectiveness and perceived quality of this toolkit. 20 users (aged 5-50 in bell curve distribution) volunteered at a local university event to test the toolkit, while conductors observed and recorded behavior from behind the exhibit. The exhibition ran for 6 hours over 2 days, and featured 2 holographic displays in a 3x3m hybrid-space. The live projected content was of a mounted, 25cm tall teddy bear (Figure 2(c)). The displays enlarged the stuffed animal to 1m tall, and ripples of data moved through the bear to make clear that the data was non-static. The two interfaces were calibrated separately to promote variable distance interaction: (1) short distance - 5-10cm from display to trigger effect, and (2) long distance - 1m from the display.



(a) (b) (c)

Figure 2: (a) Data from a model is segmented, and flies forward in virtual space. (b) User triggers isolated color flashes on model. (c) Teddybear scale reference compared with holographic display size.

4 RESULTS & DISCUSSION

User Age. It was observed that 100% of participants age 10 and below instinctively approached the holographic interfaces and attempted to interact using their hands. It was also observed that older participants were more inclined to not attempt interaction, and often did not notice that they were influencing the scene before their gaze was diverted.

[5]http://textalk.moe-nifty.com/amid/2012/05/post-66e6.html

Camera Limitations. The position of the body tracking camera in the ceiling far corner of the exhibit presented minor problems to the system, primarily tracking obstruction and distance calibration. This could be improved by mounting the Kinect closer to the displays near the floor.

Probing per Request. Participants who learned the optimal distances for interaction spent longer in the scene trying to achieve desired effects than those who quit before acclimating to the distance. One participant requested the probe to be shifted to their feet, so that they could "kick" the virtual data, which was easily configurable by our toolkit.

From our early results, we believe that initial engagement is socially influenced, thus a learned behavior. We also propose that there should exist an optimal ratio of display scale to user FOV, derived from proximity, making interactivity more visually apparent. This toolkit can help popularize the expectation of augmented space reactivity. A single camera was sufficient to perform bare-body user tracking for this interaction design; a cost effective alternative to virtual controllers for open-area interfaces. Some participants demonstrated the ability to multiplex their interaction with both interfaces simultaneously. A user feedback survey will be deployed to gauge system usability via System Usability Scale, and user enjoyment via Self-Assessment Mankikin Scale and a Frustration Discomfort Scale.

REFERENCES

Mohammad Hosseini and Christian Timmerer. 2018. Dynamic Adaptive Point Cloud Streaming. In *Proceedings of the 23rd Packet Video Workshop* (New York, NY, USA, 2018-06-12) *(PV '18)*. Association for Computing Machinery, 25–30. https://doi.org/10.1145/3210424.3210429

Mikito Takenaka, Takashi Kakue, Tomoyoshi Shimobaba, and Tomoyoshi Ito. 2021. Interactive Holographic Display for Real-Time Drawing and Erasing of 3D Point-Cloud Images With a Fingertip. *IEEE Access* 9 (2021), 36766–36774. https://doi.org/10.1109/ACCESS.2021.3062877 Conference Name: IEEE Access.

Xiao Tang, Xiaowei Hu, Chi-Wing Fu, and Daniel Cohen-Or. 2020. GrabAR: Occlusion-aware Grabbing Virtual Objects in AR. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2020-10-20) *(UIST '20)*. Association for Computing Machinery, 697–708. https://doi.org/10.1145/3379337.3415835

Qiangeng Xu, Xudong Sun, Cho-Ying Wu, Panqu Wang, and Ulrich Neumann. 2019. Grid-GCN for Fast and Scalable Point Cloud Learning. (2019), 10.

Kazuma Yoshino, Hiroyuki Kawakita, Takuya Handa, and Kensuke Hisatomi. 2020. Viewing Style of Augmented Reality/Virtual Reality Broadcast Contents while Sharing a Virtual Experience. In *26th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2020-11-01) *(VRST '20)*. Association for Computing Machinery, 1–3. https://doi.org/10.1145/3385956.3422110

Songhai Zhang, Xiangli Li, Yingtian Liu, and Hongbo Fu. 2020. Scale-aware Insertion of Virtual Objects in Monocular Videos. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 36–44. https://doi.org/10.1109/ISMAR50242.2020.00022 ISSN: 1554-7868.