HoloR: Interactive Mixed-Reality Rooms

Carsten Schwede, Thomas Hermann
Ambient Intelligence Group
CITEC, Bielefeld University, Germany
{cschwede, thermann}@techfak.uni-bielefeld.de

Abstract—**Existing virtual reality technologies conver only certain areas of the mixed-reality spectrum: Augmented reality goggles are unable to provide immersion while head-mounted displays make it difficult to interact with the real world. In this paper we introduce HoloR - short for Holographic Room: A stereoscopic, multi-person, multi-viewer, spatial projected augmented reality system, which enables applications to blend between different parts of the mixed-reality spectrum. By using web-technologies like JavaScript and WebGL the operation of HoloR does not require any installation or compiling process and thus enables rapid development of applications and extensions. We present different sample applications ranging from collaborative data exploration to augmented persons and ambient information systems.**

I. Introduction

In 1991, Mark Weiser of Xerox PARC coined the term "embodied virtuality" as an opposition to "virtual reality" which gained traction at the time. Instead of immersing people into simulated worlds he proposed to "draw computers out of their electronic shells" and make them "invisible to common awareness" [1]. Ever since, his ideas of ubiquitous computing have been influential to concepts like tangible user interfaces and smart devices as well as calm computing and ambient information systems. The recent years showed technological developments in both areas of augmented and virtual reality. On the one hand, head-mounted-displays like Oculus Rift create convincing virtual worlds but block out the real world entirely and thus make your body and those of collaborators invisible. This leads to difficulties in establishing interaction patterns that rely on spatial relationships between e.g. one's own body and virtual objects. Spatial projection displays like CAVE allow a high degree of immersion without "hiding the bodies", but also block large parts of the remaining real world with expensive, dedicated projection wall setups, which are difficult to set up. On the other hand, augmented-reality goggles like Google Glass or Microsoft HoloLens incorporate the real world, but suffer from limited field-of-view, as do smart phone-based AR applications. These technical limitations in display technologies force applications to act only on small parts of the *virtuality continuum* [2] at once (see Figure 2). For example, current technology does not enable you to start with the augmentation of a real world object (AR), use real world tools to control the virtual world (Augmented Virtuality, AV) and increase the augmentation gradually to eventually reach a degree of immersion (VR).

In this paper we introduce HoloR - short for Holographic Room: A stereoscopic, multi-person, multi-viewer, spatial projected augmented reality system that enables the rapid development of applications across the mixed-reality spectrum. Although it does not comprise a holographic display, the
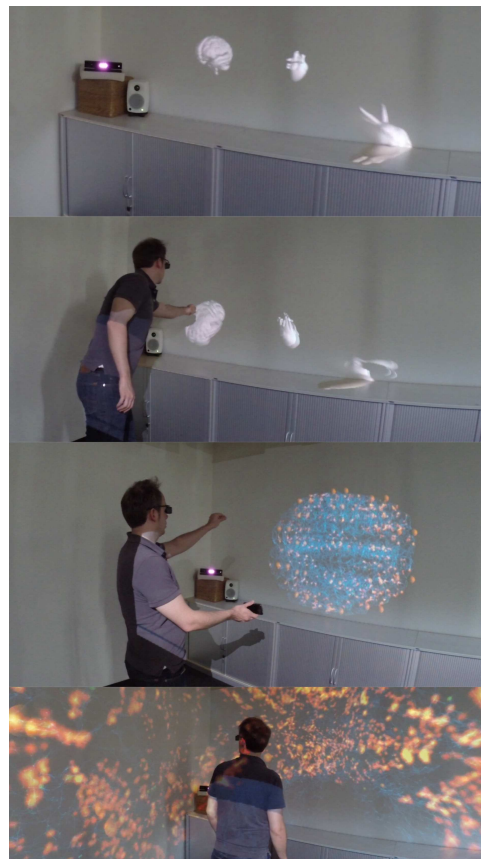

Fig. 1: Various degrees of augmentation/immersion within a single application. The brain model renders the activity of electrodes according to an EEG data set with epileptic episodes.

induced effect is often compared to holographic imaging by its first-time users and the term is useful in conveying the idea to a broader audience.

It overcomes the above limitations and enables a single application to act on multiple parts of the *virtuality continuum* or blend between them depending on the current state of interaction. For example, HoloR makes it possible, that

- the physical world is augmented with virtual objects (e.g. existing furniture has virtual objects) (AR),

- persons can interact with these virtual objects using (bi-)manual gestures as well as real world tools (AV),

- persons can zoom and step into virtual models and immerse themselves in the data (VR).
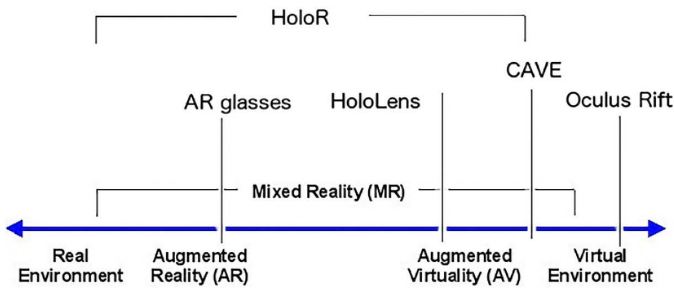
Fig. 2: Virtuality Continuum according to Milgram [2], with current systems positioned according to the author's interpretation.

A demonstration of this process is show in Fig. 1: An EEG data set with epileptic episodes is represented by a brain model floating above the sideboard and is selected by being grabbing. Hands and smart phones can be used as tools to control parameters of the virtual object. Finally, zooming and stepping into the model creates an immersive effect, despite using only a single projector.

In the following sections, we will give a condensed overview about related works and the components of HoloR. Finally we will present five sample applications in three categories that highlight different features of the system and its limitations.

Please note that most photographs in this paper are not taken from the viewers perspective and thus appear to be distorted as well as blurred due to the stereoscopic rendering.

II. RELATED WORK

The basic idea of HoloR was already mentioned in 1998, when Raskar et al. described the *office of the future*, where light bulbs are replaced with intelligent projectors and camera units [3]. This would allow the system to "autocalibrate for geometric, intensity, and resolution variations resulting from irregular [...] display surfaces". Raskar et al. further envisioned to "track a person's hands so that they could reach out and manipulate a floating 3D model" and "detect collisions between real and virtual objects so that virtual objects could be placed on the desk". In 2014, Jones et al. created units of projectors and Kinect depth cameras in the RoomAlive project and used them to project gaming content on irregular living room surfaces. Both HoloR and RoomAlive inherited the calibration and rendering process from the *office of the future* and although both share underlying techniques, they have very different goals: RoomAlive aims to create a highly immersive gaming environment for multiple players in living rooms. They sacrifice stereoscopic rendering and limit the distance of virtual objects to the projection surface so a single viewpoint of the scene can be rendered regardless of the placement and number of users. This is a reasonable thing to do as you cannot control the physical movement of the players. Limiting their area of action would have adverse effects on the gaming experience. HoloR instead aims to deliver experiences across the mixed-reality spectrum without hiding physical objects,

but instead pro-actively incorporates them into virtuality. Rendering a stereoscopic viewport for every user might limit the movement of multiple users to reasonable configurations but at the same time enables collaborative interactions in 3D. This allows users in face-to-face situations to place their hands at the same virtual point in space and interact with each other without hindering important non-verbal utterances like facial expression or gestures [4]. These "back channels" give helpful cues to avoid confusion during collaborative tasks and support alignment in communication. Blocking these may lead to "fractured interaction" and may force participants to explain their actions in detail to avoid misunderstandings [5] and hinders their coordination efforts [6]. Using physical objects or smart devices as links to the digital world are increasingly common [7], [8], [9], [10]. Jacob et al. (2008) [11] argue that successful user interfaces should not only incorporate elements from the physical realm but draw from the associated experiences and acquired skill sets in order to avoid the need to learn interface-specific skills and reduce overall mental efforts. Accordingly, every virtual object in the HoloR system has physical properties including size, mass, frictional resistance and can interact with other objects (virtual and real) through the physics engine.

III. SYSTEM OVERVIEW

HoloR consists of at least one projector and at least one tracking device. Calibration is done once using C++/OpenGL while the actual rendering is performed in a browser using JavaScript/WebGL, receiving Kinect face- and tracking results via UDP on a LAN.

A. Hardware Setup

Up to six commodity 3D short-throw projectors (in our case BenQ TH682ST with LCD shutter glasses) can be connected to a mid-grade GPU (GTX 770) in the rendering machine (i7-4790k). The number of tracking cameras (Microsoft Kinect Version 2) is not limited and can be increased to cover larger interactions volumes as well as improve the robustness of tracking. Each depth-camera is connected to an Intel NUC i5 and broadcasts its tracking data to the rendering machine via UDP. The total costs for this system are around 1.000 € for small rooms (projection area <30 square meters) and up to 8.000 € for classrooms (projection area <200 square meters).

B. Calibration and Setup

In order to transform the different coordinate systems of the various projectors and cameras, a world coordinate system needs to be established. This is done by importing a 3D model of the room created before either via CAD or Kinect Fusion [12]. Creating this model manually takes some time depending on the room's volume and complexity, but gives better projection results than hole-filling and meshing the depth image. Having an accurate ground truth model of the scenery also allows semi-automatic registrations methods that help in aligning the tracking cameras to the world coordinate system. Subsequent changes in the projection surface need to be applied to the 3D model as well to avoid distortion artefacts.

In order to estimate the intrinsic and extrinsic parameters of each projector, a gray code sequence as seen in Fig. 3 is

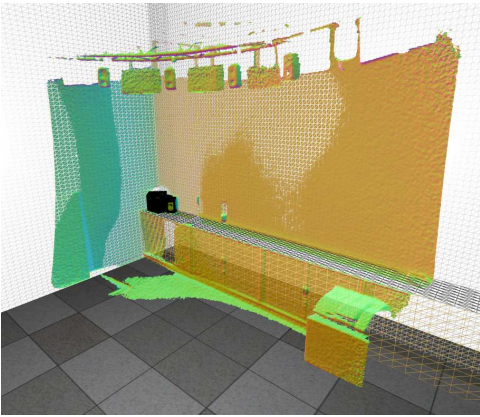Fig. 3: Projected gray code to gather correspondences to the depth camera's 3D points.

Fig. 4: Iterative Closest Point completed, Kinect point cloud aligned to 3D model of the room.
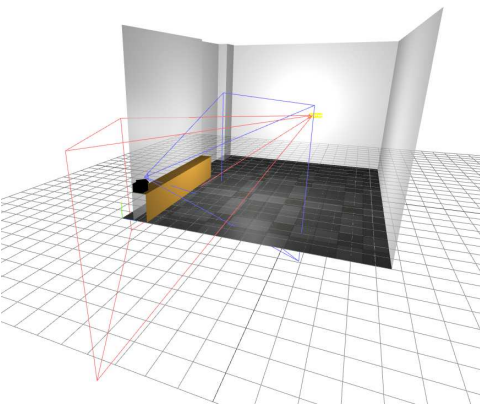

Fig. 5: Estimated frustum parameters for Kinect tracking camera (blue) and projector (red) after calibration.

displayed while the calibration software gathers pairs of the encoded 2D pixel coordinates (projector's coordinate system) and its corresponding 3D point (depth camera's coordinate system). The actual parameter estimation is performed using OpenCV's calibrateCamera method while removing outliers using a RANSAC approach.

In order to transform the extrinsic parameters of each projector with regard to the world coordinate system, the Iterative Closest Point algorithm (ICP, [13]) is used to estimate a rigid transform to map each depth camera's 3D point cloud to the 3D room model (see Fig. 4 and Fig. 5). The user may specify a number of correspondences manually by selecting them in the 3D view of HoloR's calibration tool. This estimates a rigid transform before starting ICP which better avoids local minima and leads to more robust results. Note: The same depth/tracking device can be used in succession to calibrate multiple projectors and perform the tracking later on.

C. Software Stack

Once the calibration routine is complete and the parameters of the projectors and Kinect cameras are known, HoloR can be started on the rendering machine simply by opening the corresponding web page implemented in HTML5, JavaScript and WebGL. Computations that do not directly affect the rendering process and might block the main WebGL thread are isolated in "HoloR-plugins" using HTML5-Web Workers (e.g. skeleton-tracking and gesture detection, physics engine, etc.)

Using HTML5/WebGL as the base technology has the following advantages:

- Cross-platform and cross-device compatibility (e.g. tablets like iPad work as well)

- No software dependencies except a recent browser

- Low barrier of entry for new developers

- HTML5 supports a variety of inputs (speech recognition, game pads, IMUs) and network connectivity (WebSockets, XHR, WebRTC)

- Large number of existing JavaScript/node.js modules

- Central hosting of core libraries avoids deployment issues and provides central bug-fixes

- HTML/JavaScript are easy/fast to develop (no building process)

D. Tracking and User Interaction

Every depth camera sends its tracking data via UDP to the rendering machine, where the coordinates are mapped to the world coordinate system and supplied to the main app and its plugins using event-based messaging. Average latency is around 100ms between a physical movement and the effects being rendered. Primary gestures based on the Kinect SDK are forwarded to the app as well (e.g. hands open/closed, mouth open/closed). Using these information, users are able to touch and grab virtual objects or to use gestures common to cell phone interfaces like pinch-zoom-rotate. Since the (empty) virtual space is an object by itself, users can also drag it to navigate within the virtual world.

E. Rendering

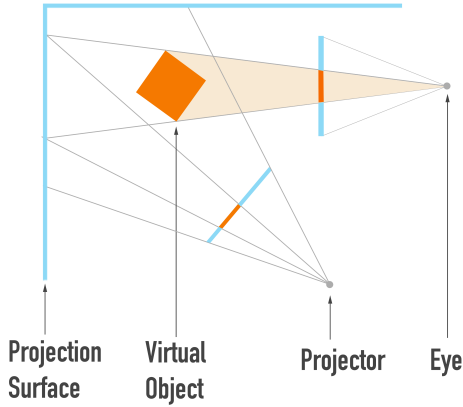The rendering as depicted in Fig. 6 consists of two passes for every frame:

Fig. 6: Schematic rendering process

1) First pass: A cube map is created by rendering the virtual world for both eyes of every person in the tracking area. The projector and the projection surface are not relevant in this pass.

2) Second pass: For every projector, a camera with identical intrinsics and extrinsics properties projects its view of the rooms projection surface geometry, ignoring the virtual world. Afterwards, a shader assigns texture coordinates for each vertex of the surface geometry by casting a ray from the vertex towards each eye while intersecting the cube map generated in the first pass. For a detailed overview of projective textures see [14].

Note that the complexity of the virtual world only affects the first pass while the complexity of the projection surface geometry only affects the second one.

This process is implemented using GLSL and is able to render 1 million sprites on our hardware in real time (60 fps) on 6 HDTV projectors in 3D for a single viewer. A single full-screen browser window spans across all connected projectors arranged in a 2×3 grid and creates a 3840×3240 WebGL context. Within this grid HoloR creates two viewports for each projector (left eye, right eye).

IV. HoloR Sample Applications

We show 5 selected applications to demonstrate the added value, here along the dimensions of collaboration, (multi-)person augmentation and ambient information systems.

A. Collaborative Data and Model Exploration

Augmented Reality has been used in educational environments ([15],[16]) as well as in collaborative working spaces [17]. We envision a Virtual Lab where machine learning tools such as 3D scatter plots or self-organizing maps are rendered as virtual tools on a physical, augmented sideboard. We expect that the physical sensation of walking (jointly with colleagues) into your data set as well as choosing tools from a physical place lowers the barrier of entry to machine learning tools and facilitates the understanding and teaching of how these tools transform or respond to a given data set. Already implemented interactions are that areas of interest in the 3D

data visualization can be focused on by using pinch-zoom-rotate gestures and by simple pointing to the relevant areas.

B. Augmented Persons

HoloR is not limited to the visualization of abstract data and models. Especially in collaborative settings, the people themselve may become a focus of interest, although not every person within the space necessarily needs to have their own view of the virtual world rendered, as portrayed in Table I. We demonstrate 3 out of the 4 cases in the following.

TABLE I: Use cases for different numbers of persons and viewers

	One Viewer	Multiple Viewers
One Person	Self-Evaluation	Presentation
Multiple Persons	Monitoring	Multi-Interaction

Monitoring data while performing actions is challenging on multiple levels: The data is usually gathered and displayed at a central point elsewhere and thus not available without disruption or distraction. At the same time, spatial context of the data is lost and embedded as an additional data point. Restoring that context and applying it back to the real world is time-consuming and prone to errors.

1) Self-Evaluation and Presentation: Real-time visual feedback can enhance motoric learning tasks, and even provide a common frame of reference for coach/student interaction, for instance in dance/sports. With our Ribbon demo we track and display the movement of a person's joints in 3D over time, much like a long-time-exposure in photography. It gives the persons and viewer(s) (which might be one and the same) an immediate feedback about the execution of the movement. Other dimensions of interest like (angular) velocity, similarity to predefined shapes could be encoded as well e.g. in terms of the ribbons colorization, width or shape.

2) Monitoring Groups: The process of monitoring patient data becomes even more challenging once multiple patients are involved. In settings like physical group therapy, the instructor has to monitor not only the execution of the movements, but the continuous well-being of the patients. Depending on the heterogeneity of the group, patients might respond differently to the same exercises or physical stress and need dedicated instructions. With HoloR, multiple persons in front of a projection surface can be monitored by a single viewer. For example, the heart rate of each person, measured by wearable health bands can be converted into a "personal physical stress value" based on the patient's prior medical assessment. This indicator is then mapped to the pulse rate, size or color of a virtual heart which is rendered atop the patients' heads. Thus HoloR provides the instructor with a valueable, real-time feedback covering the whole group at once.

3) Multi-Interaction: In order to avoid the limitations that come with rendering multiple viewports we can place the virtual objects within the projection surface. This prevents overlapping viewports and removes the need for 3D glasses (similar to the RoomAlive project). By adding projectors that project to the floor we can use the space between persons to project relationships between them, e.g. similarities within their social network profiles as seen in the bottom Fig. 7b.
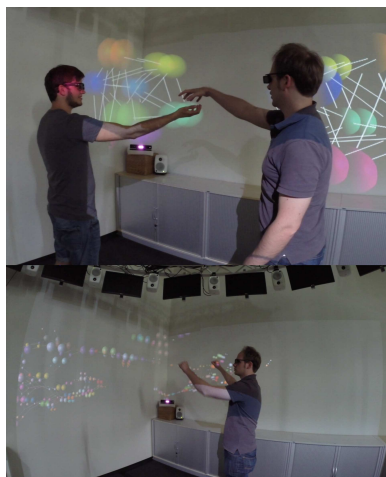
(a) Collaborative and Single User Data Exploration. **Top:** Two viewers grabbing the same sphere while both viewports are being rendered. **Bottom:** Single user graph exploration.
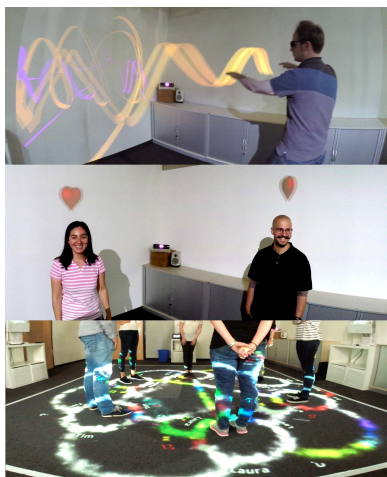
(b) Augmented Persons. **Top:** Self-evaluation of 3D motion during the Ribbon demo. **Middle:** Monitoring groups by visualizing heart rates. **Bottom:** Multi-interaction during a visualization of similarities between Facebook profiles.
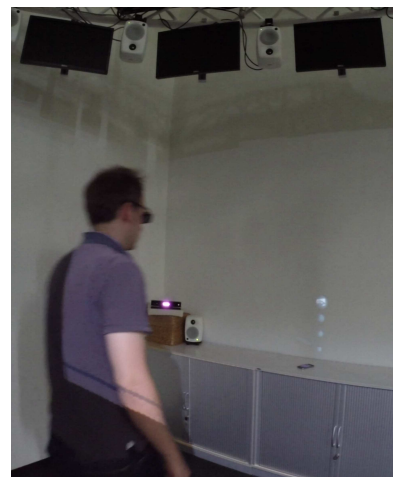
(c) Ambient Display. A notification is hovering above the smart phone lying on a sideboard.

Fig. 7: Sample Applications for HoloR.

C. Ambient Information System

Ambient Information Systems (AIS) convey non-critical information by representing them within the environment. AIS have the ability to move information "from the periphery to the focus of attention and back again" without distracting the user [18].

Since HoloR knows and projects onto large parts of the room's geometry and tracks the users' gaze direction, AIS applications can easily be developed within the HoloR framework. For instance, objects can appear at the periphery of a user's field-of-view first and slowly blend into focus (e.g. in terms of position, size, motion, opacity). Projected information could only change when the viewer is not looking in the corresponding direction in order to avoid distraction. By correlating the accelerometer data of a smart phone with those of the tracked joint positions of the user, we can estimate if and which hand is carrying the device. This allows us to estimate the absolute position of the device when being placed down and allows us later on to embed notifications above the device in 3D. Calls could be picked up or text message could be answered by gestural interaction within the 3D volume above the smart phone. From the perspective of the user the phone is transformed from being a "smart device/object" into a "hybrid object" - one that has a "normal" pattern of interaction (without augmentation) and another distinct one while being augmented. Going even further, displayed information need not be physically bound to a specific device but instead can also be displayed anywhere in the room, e.g. it could follow the user.

V. DISCUSSION

Besides creating an affordable cross-platform, multi-user mixed-reality system applicable in arbitrary furnitured rooms, HoloR promotes as a peculiar innovation the bridging of previously loosely coupled points on the virtuality continuum, and the seamless transition within a single app. Of course, the current technology comes along with certain limitations. In order to accurately reproduce the stereoscopic effect, the interpupilary distance (IPD) of every viewer has to be known. Assuming the population's mean value shows convincing effects in all single-person demos, but can lead to notable differences in collaborative scenarios, when people try to point at the same (small) object. This could be alleviated in the future by having the viewer determine her IPD through a calibration game once. Radiometric compensation is important to account for the different reflective properties of various surfaces but is limited by the maximum brightness of the projectors. This affects the maximum amount of immersion achievable with spatial projection systems in general. A good physics engine is needed to simulate convincing interactions with virtual objects and real objects or joints. As of yet it is unclear if JavaScript is capable to provide this functionality at high enough frame rates in complex scenes with multiple persons present. Multiple viewers can view virtual objects in 3D as long as the corresponding projection area of one user is in the periphery or outside of the other user's field-of-view. This constraint could be alleviated by time-multiplexing the users' viewport into the 3D shutter glasses while darkening the perceived image [4]. Other solutions like adding polarized filter or anaglyph glasses have similar drawbacks but are also limited in the maximum number of simultaneous viewers. As long as the room offers enough space, displaying multiple viewports at the same time may restrict the actors placement in some ways, but does not affect the image quality. The noise level of today's projectors limit the use of HoloR in ambient settings, but we expect ongoing developments in LED and laser projection technology to improve on this considerably.

VI. Conclusion and Future Work

HoloR offers a versatile and low-cost solution to augment common rooms while reducing infrastructure requirements and setup efforts. It can be used in educationary context in classrooms, for medical/physical monitoring, as a small CAVE alternative or as an ambient display. Depending only on a recent browser with WebGL capability, the barrier of entry for participation is very low and allows developers to quickly contribute to HoloR. Future user studies are needed to explore usability and interaction patterns in the scenarios described in the Sample Application section. We expect HoloR to perform similar to existing devices at least in the AR/AV spectrum. Because spatial projection technologies can hide reality only to a limited degree, the highest level of immersion remains in the domain of head-mounted devices. It is expected that upcoming AR goggles or lenses will at some point make spatial projection obsolete for the use cases described here. In the meantime, HoloR offers a highly apt system to develop applications at different parts of the mixed-reality spectrum and offers a test bed for current interaction patterns and those that might emerge for applications that operate on multiple areas of the virtuality spectrum at once.

Acknowledgment

References

[1] M. Weiser, "The Computer for the 21st Century," *Scientific American*, vol. 265, no. 3, pp. 94–104, 1991.

[2] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented Reality: A class of displays on the reality-virtuality continuum," *Systems Research*, vol. 2351, no. Telemanipulator and Telepresence Technologies, pp. 282–292, 1994. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.6861

[3] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays," *SIGGRAPH '98 Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 1–10, 1998.

[4] M. Agrawala, A. Beers, and I. McDowall, "The two-user Responsive Workbench: support for collaboration through individual views of a shared space," *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 327–332, 1997. [Online]. Available: http://portal.acm.org/citation.cfm?doid=258734.258875

[5] C. Heath and P. Luff, "Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction," *Human-Computer Interaction*, vol. 7, no. 3, pp. 315–346, 1992.

[6] P. Luff, C. Heath, H. Kuzuoka, J. Hindmarsh, K. Yamazaki, and S. Oyama, "Fractured Ecologies: Creating Environments for Collaboration," *Human-Computer Interaction*, vol. 18, no. 1, pp. 51–84, 2003.

[7] M. Funk, O. Korn, and A. Schmidt, "An augmented workplace for enabling user-defined tangibles," *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pp. 1285–1290, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2559206.2581142

[8] J. A. Walsh, S. von Itzstein, and B. H. Thomas, "Ephemeral Interaction Using Everyday Objects," in *Proceedings of the Fifteenth Australasian User Interface Conference - Volume 150*, ser. AUIC '14. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2014, pp. 29–37. [Online]. Available: http://dl.acm.org/citation.cfm?id=2667657.2667661

[9] C. S. Pinhanez, *The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces*. Atlanta, Georgia, USA: Springer-Verlag, 2001.

[10] V. Heun, "Smarter Objects : Using AR technology to Program Physical Objects and their Interactions."

[11] R. Jacob, R. Jacob, A. Girouard, A. Girouard, L. Hirshfield, L. Hirshfield, MS, MS, 2008, and 2008, "Reality-based interaction: a framework for post-WIMP interfaces," *Portal.Acm.Org*, pp. 201–210, 2008. [Online]. Available: http://portal.acm.org/citation.cfm?id=1357089

[12] R. A. Newcombe, D. Molyneaux, D. Kim, A. J. Davison, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-Time Dense Surface Mapping and Tracking."

[13] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, apr 1992. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/026288569290066C

[14] O. Bimber and R. Raskar, "Alternative Augmented Reality Approaches : Concepts , Techniques , and Applications," 2003.

[15] J. M. Gutiérrez, M. Dolores, M. Fernández, and J. M. Gutiérrez, "Augmented Reality Environments in Learning , Communicational and Professional Contexts in Higher Education ! Abstract ! Entornos de Realidad Aumentada aplicados al aprendizaje y a contextos comunicativos y profesionales en educación superior Resumen," no. 26, pp. 22–35.

[16] S. Cuendet, Q. Bonnard, S. Do-Lenh, and P. Dillenbourg, "Designing augmented reality for the classroom," *Computers and Education*, vol. 68, pp. 557–569, 2013.

[17] K. Arthur, T. Preston, and R. Taylor, "Designing and building the pit: a head-tracked stereo workspace for two users," *2nd International ...*, pp. 1–10, 1998. [Online]. Available: http://www.cs.unc.edu/ whitton/ExtendedCV/Papers/1998-IPT98-Arthur.pdf

[18] Z. Pousman and J. Stasko, "A taxonomy of ambient information systems," in *Proceedings of the working conference on Advanced visual interfaces - AVI '06*. New York, New York, USA: ACM Press, 2006, p. 67. [Online]. Available: http://dl.acm.org/citation.cfm?id=1133277 http://portal.acm.org/citation.cfm?doid=1133265.1133277