Avatar2Avatar: Augmenting the Mutual Visual Communication between Co-located Real and Virtual Environments

Robin Horst[1,2,3], Sebastian Alberternst[1], Jan Sutter[1], Philipp Slusallek[1], Uwe Kloos[2] and Ralf Dörner[3]
[1]*German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany*
[2]*Reutlingen University of Applied Sciences, Reutlingen, Germany*
[3]*RheinMain University of Applied Sciences, Wiesbaden, Germany*

Keywords: Mixed Reality, Video Avatar, Multi-user Environments, Low-cost, Knowledge Communication.

Abstract: Virtual Reality (VR) technology has the potential to support knowledge communication in several sectors. Still, when educators make use of immersive VR technology in favor of presenting their knowledge, their audience within the same room may not be able to see them anymore due to wearing head-mounted displays (HMDs). In this paper, we propose the Avatar2Avatar system and design, which augments the visual aspect during such a knowledge presentation. Avatar2Avatar enables users to see both a realistic representation of their respective counterpart and the virtual environment at the same time. We point out several design aspects of such a system and address design challenges and possibilities that arose during implementation. We specifically explore opportunities of a system design for integrating 2D video-avatars in existing room-scale VR setups. An additional user study indicates a positive impact concerning spatial presence when using Avatar2Avatar.

1 INTRODUCTION

One of the core aspects which connects areas of human mediated knowledge communication is the mutual communication of humans. At that, the visual aspect is the most crucial one because about 55% of a communication is transported visually (Mehrabian and Ferris, 1967).

When it comes to the usage of Virtual Environments (VEs), most VR setups constrain the communication due to wearing a closed VR HMD. This leads to a lack of a mutual representation that can be beneficial for co-located knowledge communication (Bronack et al., 2008). As a consequence of using such technology immersed learners cannot see non-immersed educators even if they are located in the same room. Thereby, the pedagogical presence, that is of importance relating to specific learning methodologies (Rodgers and Raider-Roth, 2006; Bronack et al., 2008; Anderson et al., 2001), is constrained. An example is a class-room setting, where teachers integrate HMD VR technology in their course. They cannot be seen by the learners, even though they are in the same room.

By referring to the model of visual awareness that Benford et al. (1994) proposed concerning collaborative VEs and applying it to our application space, no *focus* is provided for the immersed learners. Their *nimbus* (the space in which users can be seen by others), however, is available for non-immersed educators, so that they visually can perceive the immersed learners. But these non-immersed educators instead cannot get insights in the VE that the immersed learners act in.

In this paper, we contribute a novel collaborative Mixed Reality (MR) system and its design aspects, called Avatar2Avatar. This system indicates to equalize the described information discrepancy and to augment the mutual awareness to enhance pedagogical presence. We show the feasibility, discuss design challenges and illustrate the integration into an existing room scale VR setup. We focus on utilizing low-cost VR technology so that e.g. costly hardware or procedures as complex device-calibration are excluded in advance. We also draw conclusions about the actual presence relating to our system, measured after Schubert et al. (2001).

Although it is possible to enhances similar co-presence environments with 3D representations in real-time (e.g. (Sousa et al., 2017; Gugenheimer et al.,

Horst, R., Alberternst, S., Sutter, J., Slusallek, P., Kloos, U. and Dörner, R.
Avatar2Avatar: Augmenting the Mutual Visual Communication between Co-located Real and Virtual Environments.
DOI: 10.5220/0007311800890096
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 89-96
ISBN: 978-989-758-354-4

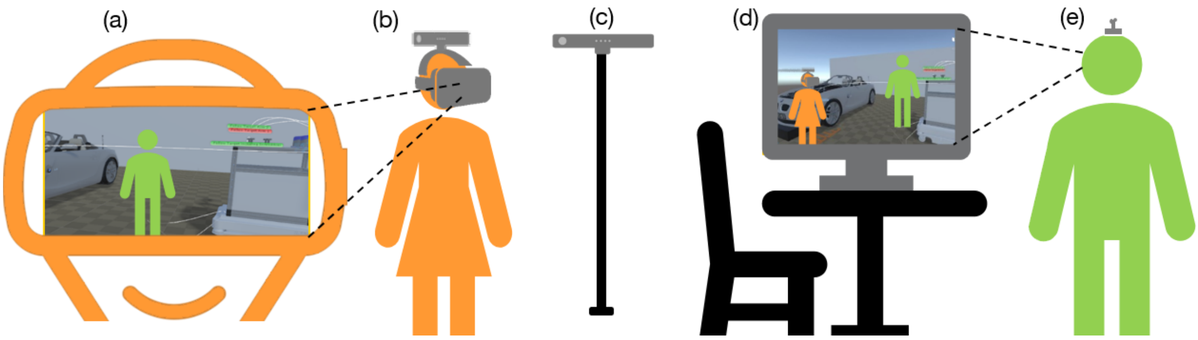

Figure 1: The setting of Avatar2Avatar: (a) virtual first-person point of view; (b) RGB-D camera attached to an immersed learner for capturing co-located educators; (c) RGB-D camera to capture all collaborators in the setting; (d) overview screen showing the virtual scene and realistic avatar representations; (e) co-located educator with tracker.

2017) we indicate that 2D video-avatar representations still provided enough visual and spatial clues to support visual communication and increase spatial presence within our learning use-case.

As the foundation of our work, we subdivide the overall system design into two sub-system concepts which work as distributed components – one for each class of users (immersed learners and non-immersed educators):

1. **Virtual POV** – The immersed learners' POV get augmented by a captured overlay image of the segmented collaborating non-immersed educators. This image is integrated into the VE as a 2D texture and then consequently rendered in the HMD (Fig. 1 (a)). A head-mounted RGB+depth camera prototype is utilized for this purpose.
2. **Scene Overview** – An overview perspective of the virtual scene and image-textures of the immersed learners and the non-immersed educators are composed and visualized on a large screen. This screen is placed within the room where both user-classes are co-located in. As a consequence, educators can orient themselves with respect to the virtual scene and visually communicate with the immersed learners (Fig. 1 (d)).

2 RELATED WORK

Pioneering work by Benford et al. (1994) and Benford and Fahlén (1993) proposes concepts and taxonomies that mostly aim for considering co-operative VEs with regard to tele-presence. They are applied in several systems. Especially in MASSIVE (Greenhalgh and Benford, 1995) the authors show how differently users interact with mutual users with respect to the avatar representation and the degree of immersion. While users with graphical representations of others keep a personal distance, users with text-only presentation of the scene are perceiving space completely different. They lack notions of this natural personal space.

Billinghurst and Kato (1999) introduce a co-located MR system that provides a three-dimensional browser. The system offers one single degree of immersion for all users and does not support immersive VEs. The presented augmented reality interface offers users to collaboratively browse web pages by using natural voice and gestures. This paper shows that with a rising degree of immersion and the tendency towards augmented virtuality (Milgram et al., 1995), the amount of visual perception of co-located users quickly decreases. The nimbus of users stays unaffected, but their focus gets diminished. Even recent system designs for co-located or tele-presence collaborations, as One Reality (Roo and Hachet, 2017), illustrate similar issues. The need of a system design that could extend existing VEs and provide mutual visual representation for users with asymmetrical different immersion is indicated here. In particular relating to multi-user environments that support more than one level of immersion this is not a trivial task, since multiple hardware setups must be incorporated in such system designs.

Our system generally relates to realistic avatar representations. There already exist several concepts and systems that deal with realistic human avatars. Each of them uses a different approach due to their area of application. Huang and Kallmann (2009) propose a system which focuses on realistic motions for avatars. Lok et al. (2014) propose a system that creates the realism aspect of avatars through high physicality. There also exists work (Kotranza et al., 2009) which introduces a system to augment an immersive VE by providing haptic aspects. Here, non-verbal communication benefits from the realistic tangibility of a virtual human avatar. In contrast to the above we integrate the realism-aspect of our avatars through real-time captured video textures of the co-present

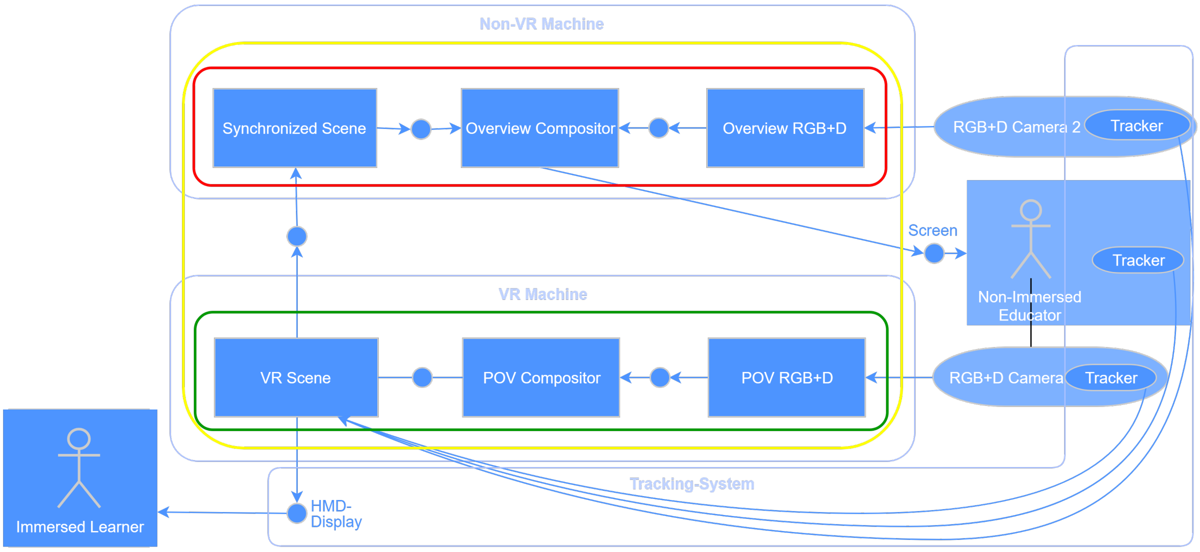

Figure 2: Compositional structure of the overall system design. The yellow shape depicts the system's boundary, the red and the green respectively a sub-system.

persons.

Current VR systems for collaborative environments similar to ours (Sousa et al., 2017; Gugenheimer et al., 2017) propose processes to integrate real-time 3D aspects, as well. Gugenheimer et al. (2017) on the one hand do not focus on avatars, but on the interactions that are provided for the different user-classes. Sousa et al. (2017) on the other hand show opportunities and use-cases for both 3D and 2D representations from video-streams for telecommunication purposes. In the latter, a toolkit is proposed that models displays as planar rectangular static surfaces within the virtual room. Our work, however, focuses on these limitations of the 2D avatar integrations in VEs and proposes a system design for flexible and non-static video-avatars.

Further work (Garau et al., 2003; Latoschik et al., 2017; Roth et al., 2016) points out the needs for nonverbal communication benefits for realistic avatars. Therefore, we refer to work considering low-cost technologies, like Kinect cameras or HTC Vive systems, which capture video textures and integrate them into VEs (Slater et al., 2010; Nordby et al., 2016; Regenbrecht et al., 2017; Zhu et al., 2016; Lee et al., 2006; Barmpoutis, 2013). Work by Lee et al. (2006) which is, however, comparable to ours integrates 2D avatars visible for only one class of users (the non-immersed educators). The integration of such avatars into the immersive part of the system (immersed learners) or into an existing VR system is not mentioned in these examples.

3 AVATAR2AVATAR SYSTEM DESIGN AND PROTOTYPE

Our system is used to visualize realistic avatar representations of all users and the VE (Fig. 1 (d)) so that collaborating non-immersed educators can see it. Simultaneously, pictorial representations of non-immersed educators are integrated in the immersive VE. The immersed learners therefore are enabled to perceive their collaborators while wearing the closed HMD (Fig. 1 (a)).

The Avatar2Avatar system design is compositionally modeled in Fig. 2 using fundamental modeling concepts (Knöpfel et al., 2005). Here, we see how the overall system (yellow boundary) is designed to be run on two machines – each with a sub-system (green and red boundary) which communicate over a local network. The system is built from six components (Fig. 2) which are described in detail in the next sections:

1. *Synchronized Scene*
2. *Overview RGB+D*
3. *Overview Compositor*
4. *VR Scene*
5. *POV RGB+D*
6. *POV Compositor*

All of the above modules can be classified into either *capturing the virtual scene* (components 1 and 4), *extracting people's image* (components 2 and 5) or *composing these with the virtual scene* (components 3 and 6).

3.1 Capturing the Virtual Scene

We utilize compositing techniques to create two separate textures: On the one hand, a full screen texture of user representations and the virtual scene, to be drawn on the external overview screen and on the other hand a texture of non-immersed educators to be integrated into the VE for the immersed learners. Therefore, we incorporate color and depth information from the VE. The modules *VR Scene* and *Synchronized Scene* (2) provide this image-based data. To capture the necessary information from the virtual scenes we point out three tasks to be performed:

Camera Alignment – Since we compose real and virtual footage, there must be an alignment of the real and the virtual camera. For matching the cameras' rotation and position we use tracking provided by the tracking system that is part of an HTC Vive VR setup. Other parameters are approximated by adjusting the virtual camera so that no further optical calibration is needed for the physical counterpart. Examples are the radial distortion of the optical lens and extrinsic/intrinsic camera parameters (like the field of view).

Color Image Acquisition – While we propose a design that is based on low-cost technologies, we make use of the game engine Unity, which is appropriate for the concept. A major challenge when acquiring visual information from game engines is the interference with the game loop. In this context, an asynchronous real time texture read-back from the already frequented GPU is compulsory. The rendering must be performed independently of the game loop so that a continuous and consistent image stream can be assured.

Depth Acquisition – In contrast to simple color image acquisition using virtual cameras, the function of acquiring depth information must be provided, too. To solve this issue a custom shader can be used to write depth information into a separate texture. This texture can asynchronously be read from the GPU the same way the color image texture is read back. The game loop which runs the existing VE remains unaffected by this integration into our system design.

3.2 Capturing and Extracting Peoples' Textures

To augment the mutual visual communication between real and virtual environments we propose to integrate 2D images of peoples as video textures into VEs that have different degree of immersion. For providing these realistic avatars we rely on real texture representations which must be extracted from camera-captured 2D images beforehand. This functionality is provided by the *Overview RGB+D* and the *POV RGB+D* modules (2).

For capturing RGB+D image resources, two Microsoft Kinects (1920x1080 RGB; 512x424 D; 30 fps) are utilized in the proposed system. A major difference in terms of processing captured images for either the *overview screen* or the *virtual POV* is the segmentation of the peoples' textures. For the non-moving camera which captures the whole scene (Fig. 1 (c) and Fig. 2 RGB+D Camera 2) the Kinect API's functionality can be used for segmentation.

The process that provides the textures for integration into the VE in contrast cannot rely on such APIs' functionality. None offers the extraction of people in images for a constantly moving camera, as it is for simulating a *first person POV*. This segmentation therefore is handled during the compositing step itself.

3.3 Composing People and the Virtual Scene

For compositing we differentiate between two modules, the *Overview Compositor* and the *POV Compositor*. They are responsible for composing the image for the *overview screen* and the *virtual first person POV* representation respectively. Both compositor modules receive similar data from the Kinects and the virtual scene to perform a depth compositing in a similar way as proposed by Zhu et al. (2016). Since there already exists work which proposes straightforward solutions to use consumer-oriented VR technology to visualize immersed people and their VE on an external screen (Zhu et al., 2016), we will focus on the novel aspects of composing video textures into immersive VEs:

Integrating Two Compositing Systems – Our system integrates the two compositing modules similar to a client-server architecture. The final image for the external screen is processed and rendered directly on the external screen. The texture for the VE in contrast must be processed in the compositor and then sent back to Unity, where additional compositing steps are processed.

Extracting People based on Room-scale VR Technology – To handle the first-person POV simulation challenge, mentioned in the previous section, we cannot use a common method as for the *overview screen* (e.g. provided by the Kinect API). In Avatar2Avatar we propose to segment them by combining a virtual bounding box (BB) (Fig. 3 (a)). The depth information is furthermore used to perform a first occlusion calculation by testing the depth

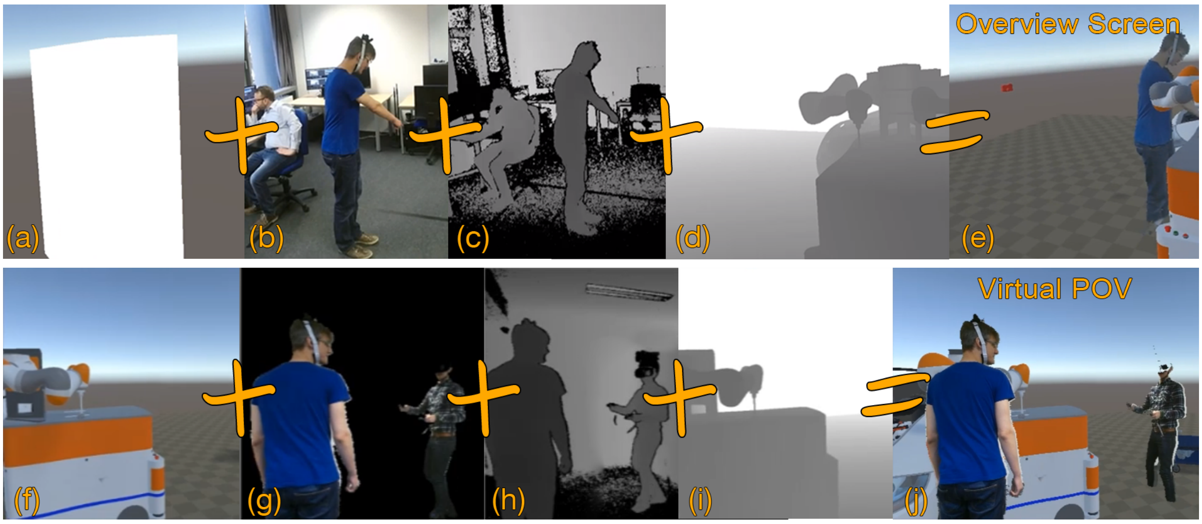

Figure 3: Footage (a)-(d) for the compositor of the *virtual POV* (e) and footage (f)-(i) for the compositor of the *overview screen* (j). a) shows the virtual bounding box of the educator (blue shirt). It is used with the color image (b), the corresponding depth image (c) and the depth of the scene (d) to compose the texture of the educator into the view of the immersed learner (checked shirt; (e)). A Kinect segmentation work-flow is incorporated in our system ((g), (h)) to compose textures of both persons with the virtual scene ((f), (i)) and render them together on an external screen (j).

of virtual objects to the captured depth of the non-immersed educators.

Composing Realistic Avatar Textures Into a VE – While the avatar texture here is already segmented and occlusions with virtual objects are realistically calculated (for counteracting false positive occlusion, Tab. 1), it still must be integrated into the VE. Therefore, we draw a 2D plane with dimensions of the texture at the position of the non-immersed educator and apply the texture to it. As we capture the texture from a first person POV of the immersed learner, we must ensure that this texture is projected on a plane orthogonally to the virtual viewing direction, as well.

Another challenge is the projection of a three-dimensional human body onto a single depth value. We propose to track the non-immersed educator so that the avatar texture plane can be positioned in the corresponding position within the VE. Since we only tested for occlusions from virtual objects it is hypothetically possible that extremities of the learners could reach out from this single depth point. These extremities could then the other way around occlude a virtual object instead. This cannot be processed beforehand due to the three-dimensionality of the virtual scene. As a consequence, we propose to offset the plane about the length of half an arm span towards the immersed user. This is a trade-off between the correct placement of the texture and what we refer to as false negative texture occlusion (Tab. 1).

Table 1: Classification of the occlusion problem into false/true negative/positive.

	Pixel is in front of the object	Pixel is behind the object
Pixel is drawn in front of the object	true positive	false positive
Pixel is drawn behind the object	false negative	true negative

4 EVALUATION

Compared to absence of mutual visual representation, related work indicates that an existing and realistic representation will improve the mutual visual communication (see section 2. Since we integrate existing VEs into our system, the impact of our concept on the presence of immersed users is of interest. A negative impact could negate the advantage of the augmented communication at the expense of the presence.

The use-case of our study was set within a collaborative training scenario, where a non-immersed trainer had to familiarize a trainee with a construction environment in the automotive section. New construction procedures were to explain. A collaboration with a robot assistant should be utilized by the immersed trainee to solve the construction task.

4.1 User Study

The study involved 12 paid, voluntary **participants** (7 male, aged 23 to 35 with Ø 26,5 and SD 3,34). The **procedure** was based on the approaches in (MacKenzie, 2012). Participants were welcomed, filled out demographics, then were asked to interactively explore the VE in cooperation with the experimenter and finally were interviewed and filled out a post-study questionnaire.

The **design** of the study included a random distribution of the participants into two groups (between-group design) - the experimental group which used the proposed enhancement of the visual communication and the control group which only used common VR technology (an unextended HTC Vive setup). The presence of the participants was measured for being the dependent variable. The IGROUP Presence Questionnaire (IPQ[1]) was used. It subdivides the presence into three units: spatial presence, involvement and experienced realism.

For analyzing the **results** of the experiment, we conducted a two sample t-test, by assuming normal-distribution on the data. Tests on separate aspects of presence revealed a phenomenon concerning the spatial presence. Three out of five questions regarding the spatial presence factor of the IPQ show a significant difference in favor to the score of the experimental group for questions Sp2 (p-value ¡ 0,00001) and Sp3 (p-value = 0,021028) with p ¡ 0,05 and Sp5 (p-value = 0,076108) with p ¡ 0,10 (Fig. 4).

Observations and user statements have indicated four phenomena: Five out of the six experimental group participants mentioned that the Kinect rig was uncomfortable to wear, particularly due to the weight of the rig which was mounted at the head. Two participants mentioned the infrequent cut off of the extremities of the avatars. One participant unexpectedly signaled that it would be comfortable if the video avatar visualization could be turned on and off by him- or herself, depending whether help was needed or not. All six participants of the experimental group explicitly expressed that they perceived the video avatar representation of the co-located educator helpful.

4.2 Discussion

The results indicate an improvement of the spatial presence of the participants that were in the experimental group. Most critical of the qualitatively measured issues (cut off limbs and rig weight) can be attributed to the specifics of the Kinect camera. This is a critical aspect of the design, because other cameras that deliver similar content (matched color and depth content) have as well similar specifications. Since we target a low-cost concept we exclude to use professional hardware to solve this challenge. A software-based solution for addressing the differing aspect ratio and resolution of the Kinect and the HTC Vive, for example as a separate module within the system design that handles multiple cameras, however, is appropriate. The camera itself could alternatively be mounted at the torso of participants so that the weight of the camera will not be perceived as focused as on the head.

As we chose to represent users within the 3D VE as a 2D textured plane, there surely are alternative representations that could be of significance. The study, however, indicates, that 2D video avatars offered enough visual and spatial clues for all of our participants to perceive corresponding information that is necessary for mutual visual communication. Since one participant remarked that the optionality of the video avatar visualization would be helpful, this phenomenon should as well be considered in future system designs. Questionnaire comments and qualitative observations during the experiment indicate that overall, participants were satisfied with Avatar2Avatar and found it particularly helpful for the given collaborative knowledge communication task.

During the implementation of the system there appeared several issues that had to be solved. Since the camera that simulates the POV is attached above the real POV (eyes) of the user, there are challenges of transforming the virtual human texture/objects into the correct angle and position. This is negligible if objects are farther away but attract more visual attention from close up. Furthermore, there exist obvious visual quality issues, like artifacts, in the avatar texture (Fig. 3 (e) and (j)). These can be attributed to the relatively low depth resolution of the Kinect, which is only 512x424 pixels. While the majority of artifacts could be eliminated using dilate and erode algorithms for edge cleaning, the artifacts are accepted in favor to ensure a stable 30 frames per second rate and low latency of the compositors. Some artifacts also exist because the statistical model of the Kinect API's human recognition is not arranged to recognize users with an HMD or even a Kinect rig on the head.

Another issue that exists due to Kinect specifics is that the virtual POV augmentation cannot cover the entire resolution of the HMD. It is calibrated to a part in the upper middle of the view. Dependent on the position, extremities of persons could be cut off the texture. The reason for this is the different aspect ratio of the Kinect image (1920x1080 pixels $\sim$ 16:9)

[1]http://www.igroup.org/pq/ipq/index.php (June 18, 2018)

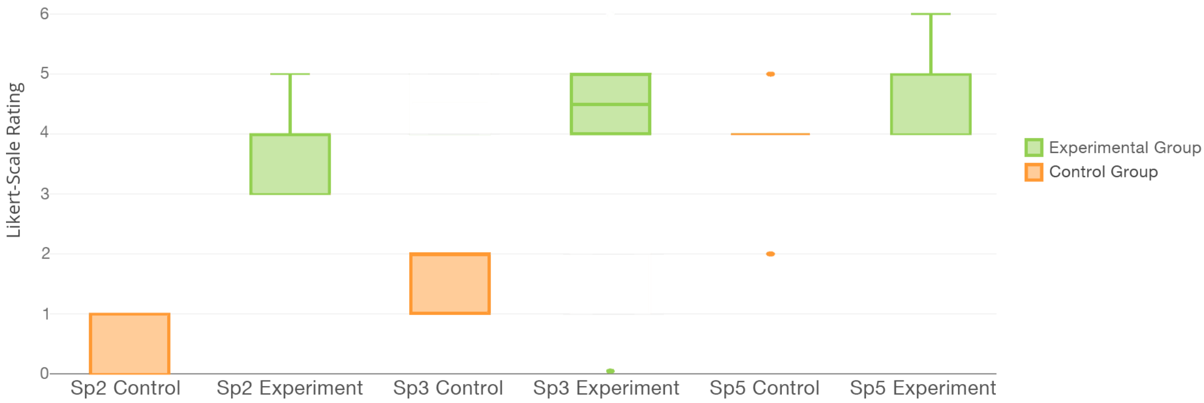

Figure 4: Three items of the IPQ show significant difference in favor to score of the experimental group.

and the Vive HMD (one eye with 1080x1200 pixels $\sim$ 9:10). As usual, a small part is not at all in the view of a single eye due to parallax of two eyes.

Finally, jittering of the virtual bounding box can be seen which results in partially cutting body parts of the human texture within the VE. This jittering occurs due to interfering infrared rays of the two Kinects, since they use the same wave length as the HTC Vive Lighthouses. Depending on the distance to each other and the orientation, more or less jittering appears.

5 CONCLUSION AND FUTURE WORK

In this paper we proposed the Avatar2Avatar system and design, which augments the mutual visual communication between co-located real and virtual environments. Non-immersed educators are involved in the VE by providing an overview about the virtual scene and all users of the system, including themselves. These non-immersed educators therefore get immersed on a low level which creates a base context for mutual awareness. In terms of the visual awareness model we provide a focus into the real environment for immersed learners that utilizes the nimbus (Benford et al., 1994) – as mentioned in section 1 – of co-located non-immersed educators. Educators get their already existing focus widened so that they can simultaneously see both the real persons and the virtual environment.

While we focused our work specifically on knowledge communication between users within one physical room, it is of major interest for us to transfer our system design to the field of telepresence. Therefore, there will be several changes necessary regarding the hardware. Especially the simulation of the dynamic virtual first person POV will be challenging. Further calibration of all devices would be necessary (as e.g. in (Beck and Froehlich, 2017)), but which is at the expense of the low-cost and consumer-oriented aspect of the setup. In the second room there must be a hardware construction that moves the camera accordingly to the immersed users POV in the first room and vice versa. Prototyping such a construction and furthermore restricting the set-up to low-cost consumer hardware could have a significant impact on knowledge communication sectors, like e-Learning or distance learning.

ACKNOWLEDGEMENTS

The work is supported by the Federal Ministry of Education and Research of Germany in the project HybriT (Funding number:01IS16026A). The work furthermore is minor supported by the Federal Ministry of Education and Research of Germany in the project Innovative Hochschule (Funding number: 03IHS071).

REFERENCES

Anderson, T., Liam, R., Garrison, D. R., and Archer, W. (2001). Assessing teaching presence in a computer conferencing context.

Barmpoutis, A. (2013). Tensor body: Real-time reconstruction of the human body and avatar synthesis from rgb-d. *IEEE transactions on cybernetics*, 43(5):1347–1356.

Beck, S. and Froehlich, B. (2017). Sweeping-based volumetric calibration and registration of multiple rgbd-sensors for 3d capturing systems. In *Virtual Reality (VR), 2017 IEEE*, pages 167–176. IEEE.

Benford, S., Bowers, J., Fahlén, L. E., and Greenhalgh, C. (1994). Managing mutual awareness in collaborative virtual environments. In *Virtual Reality Software and Technology*, pages 223–236. World Scientific.

Benford, S. and Fahlén, L. (1993). A spatial model of interaction in large virtual environments. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW93*, pages 109–124. Springer.

Billinghurst, M. and Kato, H. (1999). Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*, pages 261–284.

Bronack, S., Sanders, R., Cheney, A., Riedl, R., Tashner, J., and Matzen, N. (2008). Presence pedagogy: Teaching and learning in a 3d virtual immersive world. *International journal of teaching and learning in higher education*, 20(1):59–69.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536. ACM.

Greenhalgh, C. and Benford, S. (1995). Massive: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3):239–261.

Gugenheimer, J., Stemasov, E., Frommel, J., and Rukzio, E. (2017). Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4021–4033. ACM.

Huang, Y. and Kallmann, M. (2009). Interactive demonstration of pointing gestures for virtual trainers. In *International Conference on Human-Computer Interaction*, pages 178–187. Springer.

Knöpfel, A., Gröne, B., and Tabeling, P. (2005). Fundamental modeling concepts. *Effective Communication of IT Systems, England*.

Kotranza, A., Lok, B., Pugh, C. M., and Lind, D. S. (2009). Virtual humans that touch back: enhancing nonverbal communication with virtual humans through bidirectional touch. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 175–178. IEEE.

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., and Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, page 39. ACM.

Lee, S.-Y., Ahn, S. C., Kim, H.-G., and Lim, M. (2006). Real-time 3d video avatar in mixed reality: An implementation for immersive telecommunication. *Simulation & gaming*, 37(4):491–506.

Lok, B., Chuah, J. H., Robb, A., Cordar, A., Lampotang, S., Wendling, A., and White, C. (2014). Mixed-reality humans for team training. *IEEE Computer Graphics and Applications*, 34(3):72–75.

MacKenzie, I. S. (2012). *Human-computer interaction: An empirical research perspective*. Newnes.

Mehrabian, A. and Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of consulting psychology*, 31(3):248.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–293. International Society for Optics and Photonics.

Nordby, K., Gernez, E., and Børresen, S. (2016). Efficient use of virtual and mixed reality in conceptual design of maritime work places. In *15th International Conference on Computer and IT Applications in the Maritime Industries-COMPIT'16*.

Regenbrecht, H., Meng, K., Reepen, A., Beck, S., and Langlotz, T. (2017). Mixed voxel reality: Presence and embodiment in low fidelity, visually coherent, mixed reality environments. In *Mixed and Augmented Reality (ISMAR), 2017 IEEE International Symposium on*, pages 90–99. IEEE.

Rodgers, C. R. and Raider-Roth, M. B. (2006). Presence in teaching. *Teachers and Teaching: theory and practice*, 12(3):265–287.

Roo, J. S. and Hachet, M. (2017). One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 787–795. ACM.

Roth, D., Lugrin, J.-L., Galakhov, D., Hofmann, A., Bente, G., Latoschik, M. E., and Fuhrmann, A. (2016). Avatar realism and social interaction quality in virtual reality. In *Virtual Reality (VR), 2016 IEEE*, pages 277–278. IEEE.

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281.

Slater, M., Spanlang, B., Sanchez-Vives, M. V., and Blanke, O. (2010). First person experience of body transfer in virtual reality. *PloS one*, 5(5):e10564.

Sousa, M., Mendes, D., Anjos, R. K. D., Medeiros, D., Ferreira, A., Raposo, A., Pereira, J. M., and Jorge, J. (2017). Creepy tracker toolkit for context-aware interfaces. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, pages 191–200. ACM.

Zhu, Y., Zhu, K., Fu, Q., Chen, X., Gong, H., and Yu, J. (2016). Save: shared augmented virtual environment for real-time mixed reality applications. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1*, pages 13–21. ACM.