HAL
open science

Vishnu: Virtual Immersive Support for HelpiNg Users - An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality

Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan, Bruno Arnaldi

To cite this version:

Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan, Bruno Arnaldi. Vishnu: Virtual Immersive Support for HelpiNg Users - An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality. 3DCVE 2016 : International Workshop on Collaborative Virtual Environments, Mar 2016, Greenville, États-Unis. pp.1 - 5, 10.1109/3DCVE.2016.7563559 . hal-01293435

HAL Id: hal-01293435

https://hal.science/hal-01293435

Submitted on 24 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vishnu: Virtual Immersive Support for HelpiNg Users

An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality

Morgan Le Chénéchal*
IRT b<>com

Thierry Duval†
Télécom Bretagne
Lab-STICC
IRT b<>com

Valérie Gouranton‡
INSA Rennes
Irisa/INRIA
IRT b<>com

Jérôme Royan*
IRT b<>com

Bruno Arnaldi‡
INSA Rennes
Irisa/INRIA
IRT b<>com

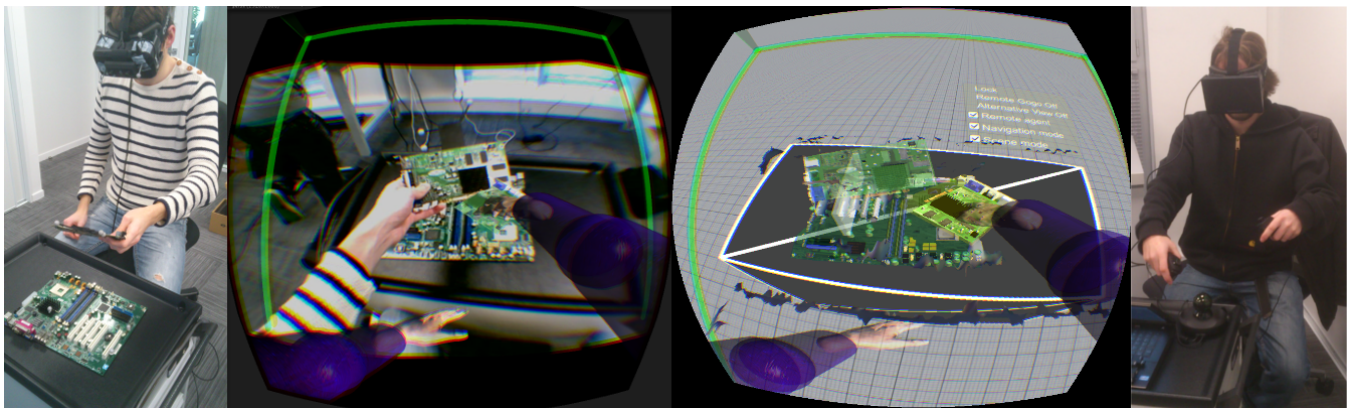


Figure 1: Illustration of the system and the viewpoints of the agent (left) and the expert (right) in a motherboard assembly scenario.

Abstract

Increasing networking performances as well as the emergence of Mixed Reality (MR) technologies make possible providing advanced interfaces to improve remote collaboration. In this paper, we present our novel interaction paradigm called *Vishnu* that aims to ease collaborative remote guiding. We focus on collaborative remote maintenance as an illustrative use case. It relies on an expert immersed in Virtual Reality (VR) in the remote workspace of a local agent helped through an Augmented Reality (AR) interface. The main idea of the *Vishnu* paradigm is to provide the local agent with two additional virtual arms controlled by the remote expert who can use them as interactive guidance tools. Many challenges come with this: collocation, inverse kinematics (IK), the perception of the remote collaborator and gestures coordination. *Vishnu* aims to enhance the maintenance procedure thanks to a remote expert who can show to the local agent the exact gestures and actions to perform. Our pilot user study shows that it may decrease the cognitive load compared to a usual approach based on the mapping of 2D and de-localized informations, and it could be used by agents in order to perform specific procedures without needing to have an available local expert.

Index Terms: H.5.1 [Information Interfaces and Presentation (e.g. HCI]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation (e.g. HCI]: User Interfaces—Prototyping; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques

*e-mail: {morgan.lechenechal, jerome.royan}@b-com.com
†e-mail: thierry.duval@telecom-bretagne.eu
‡e-mail: {valerie.gouranton, bruno.arnaldi}@irisa.fr

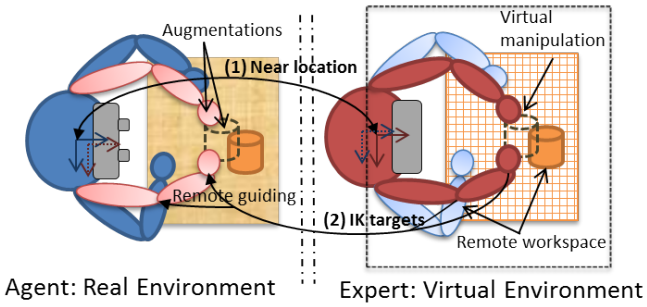


Figure 2: A conceptual top view of the *Vishnu* paradigm.

1 Introduction

Remote guiding of an agent performing a physical task helped by an expert has already proved its usefulness in many proposed systems [11, 6]. It can be used to achieve punctual unknown procedures, decreasing the time and the cost of the interventions and allowing a single expert staying in the same place to help several potentially dispersed remote agents. To achieve such remote guiding with the same efficiency as a local expert, the proposed system must provide some essential functionalities:

- The agent must be able to perform free hand-gestures, and to perceive augmentations integrated to the real space controlled by the expert, using a head-worn display.
- The remote guiding gestures must be easy to map in the agent's real 3D world.
- The expert must be collocated in the distant workspace in order to have a proper viewpoint on it regarding the scale of the scene and to ease his interactions.
- The expert must be able to interact with virtual objects that must be seen by the remote agent.

Vishnu is a novel interaction paradigm that proposes to add two virtual arms to a local agent controlled by a remote expert. With these two additional arms coming out of the agent's own shoulders, the remote expert can use them as interaction tools and show the exact gestures and actions to perform, as illustrated in Figure 2. Therefore, interaction is more natural for the expert, and it is easier for the agent to understand the remote guiding instructions.

Section 2 introduces related work and similar systems proposed in the literature. Then, we present the *Vishnu* paradigm in Section 3 and our implemented setup in Section 4. Last, Section 5 evaluates our technique with a pilot user study, and Section 6 concludes this paper and proposes perspectives on this work.

2 Related Work

In this section, we do not refer to work that focuses on Hand-Worn Display (HWD), because we want to let the agent interacts with free hands gestures.

Advanced telepresence systems have been developed and propose to share a distant face-to-face meeting [14], taking advantages of depth cameras to reconstruct a distant place. The shared virtual place created erases the real distance between collaborators, and some systems even provide interactions around a common device or a touch table. Others focus on local collaboration and propose to share a virtual and augmented environment in the same time through a Mixed Reality (MR) system [8, 3]. Users can seamlessly switch between a Virtual Reality (VR) or an Augmented Reality (AR) setting in order to take advantage of multi-scale viewpoints. However, these systems do not provide collaborative features able to handle the guidance of a remote expert in performing a physical task.

Conversely, Alem et al. developed many MR interfaces for remote collaborative maintenance. HandInAir [5] and MobileHelper [11] are examples of systems that provide gesture-based guiding for an agent helped by a remote expert. The agent wears a helmet with a camera that is streamed to the helper. Then, the expert acts with free mid-air gestures in front of a camera, and his hands are merged on the output display of both users. Extensions using 3D capturing have been proposed with HandsIn3D [7] and the 3D Helping Hands [12] systems. They add 3D handling of occlusions and virtual shadows and, thus, improve immersion and the sense of presence. Nevertheless, the displays are still not truly collocated with the real environment (because of the use of a near-eye display and no tracking), and the interaction of the helper stays limited to moving his head around the position of the agent's head and to showing the gestures to perform without being able to interact with any object.

3 The Vishnu Paradigm

The main idea of *Vishnu* is to add two virtual arms to an agent, coming out of his own shoulders. Thus, the agent can see his two own real arms in addition to the two additional virtual arms that are controlled by a remote expert, who can use them as interactive guidance tools. Figure 1 illustrates this paradigm that could be used in many applications: industrial maintenance, procedure learning and sports training. It is based on a bi-directional communication (cf Fig. 2). The expert's virtual location must be close to the agent's real position and orientation (1). The way the expert navigates in the shared VE is described in subsection 3.2. Then, the expert's virtual arm gestures define Inverse Kinematics (IK) targets used to control the virtual arms of *Vishnu* for the agent (2). With *Vishnu*, we overcome the limitations highlighted in the related work for both the expert and the agent in order to improve their sense of co-presence and to decrease their cognitive load due to a de-localized display and interactions.

In the following, we present the features of our system, including the *Vishnu* paradigm in addition to others that provide a highly usable system.

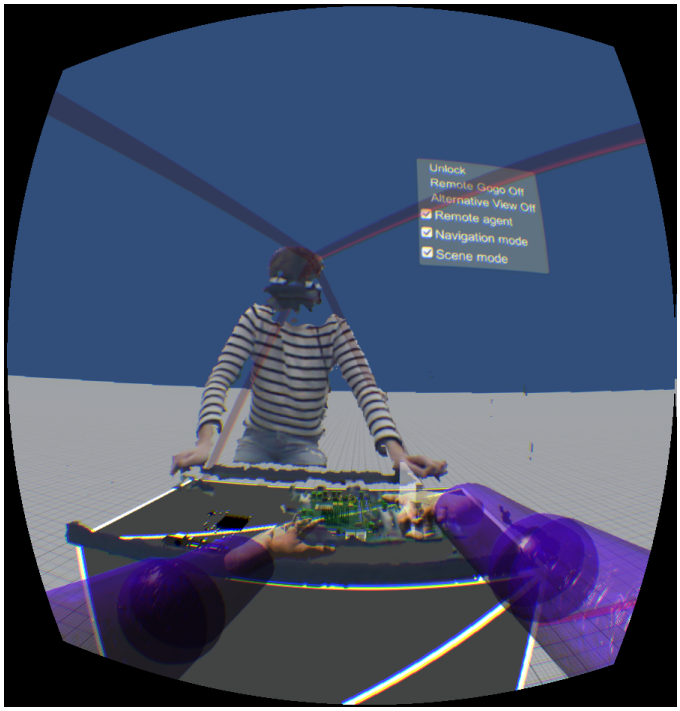


Figure 3: An illustration of the remote expert's viewpoint standing in front of the distant agent 3D reconstruction.

3.1 Agent Features

First, a main goal of our system is to provide the agent with a true AR display enabling a proper perception of the augmented environment. To achieve this, we need to dynamically reconstruct the real environment in 3D in order to handle occlusions of real objects by virtual ones (cf. section 4 for implementation details). Moreover, this allows the handling of virtual shadows that increase the sense of presence and ease the perception of depth. Furthermore, the agent's head is tracked to provide a collocated stereoscopic display.

The *Vishnu* paradigm is based on an IK algorithm [13] with virtual targets for the elbows and hands while the shoulders are fixed regarding the agent's ones. This makes sense only if the agent is close to the expert's virtual location. Otherwise, the virtual targets are too far from the hands and elbows controlled by the expert, and the remote gestures are not relevant anymore. Thus, the expert is represented with his viewing frustum that changes its color from red to green according to the distance between the two users. This provides basic feedback to the agent that enables him to place himself correctly when necessary. In the same way, the virtual arms smoothly disappear when the agent moves away from the expert's location.

3.2 Expert Features

The strongest limitation found in previous work is the limited interactions of the expert. Our system provides a VR setup for the expert that enables him to interact in the shared VE in a richer way. In addition to the control of his viewpoint, the expert controls two virtual arms that can interact with virtual objects. Thus, he can grasp and manipulate virtual objects that are displayed as augmentations on the agent side (cf Fig. 1).

The expert's arm parts (i.e. shoulders, elbows and hands) are tracked and used to control their virtual representations for proper kinematics between the expert's gestures and his interactive tools.

The manipulation of a virtual object is achieved using an interaction technique based on a virtual hand metaphor [2].

Concerning navigation, the expert can freely move in the VE. But, to guide the remote agent with his gestures, the expert must be close to the agent's location. Two interactions are possible:

- Smoothly bringing back his location and orientation to the remote agent's current one.
- Locking/unlocking the automatic following of the remote agent's head by the expert's head.

When the expert shares the viewpoint of the agent, a navigation technique based on the famous GoGo arms can be enabled [10]. This technique allows the expert to guide the agent, using the *Vishnu*'s arms, in order to place himself correctly to begin the manipulation task.

Last, the 3D reconstruction of the remote agent and workspace can be enabled/disabled depending on current awareness needs (cf Fig. 3). Especially in the diagnostic phase, the expert can stand face to the agent and freely explore the remote workspace in order to find the correct place to perform the maintenance task.

4 Implementation

For the agent, the 6 Dof tracking of his head and of real objects (initially overlaid with virtual clones) is performed with an OptiTrack™V120:Trio IR tracking system. The 3D reconstruction is based on Kinect™. The AR display is a video see-through HMD made with an Oculus Rift™equipped with two front stereo cameras in order to provide AR with wide field of view.

For the expert, the current system uses Razer™Hydra. Using one controller for each hand allows a 6 Dof tracking and provides buttons to interact and joysticks to navigate. The tracking of the expert's shoulders and elbows is based on Kinect™. Last, the display can be done with a desktop screen or with an HMD such as Oculus Rift™to handle head-tracked 3D stereoscopy.

Applications are based on Unity3D™C# scripting, and CG shaders handling Kinect-based real-time 3D reconstruction (also used for AR virtual occlusions and shadows). In future versions of our system, this 3D reconstruction of the agent's workspace should be done based on the stereo cameras, and the optical tracking system replaced by image-based tracking (i.e SLAM and/or 3D models-based). Moreover, the expert's VE should not need a pre-modeling phase, and could be dynamically generated based on the 3D reconstruction coupled with interactive meshes segmentation and fusion [4].

5 Pilot user study

We ran a pilot user study with few subjects that were asked to compare our system (mode 1) versus a basic one (mode 2) based on a desktop screen on the agent side and a fix camera streamed to the remote expert; from this point of view, the expert could sketch in 2D on the camera stream in order to guide the agent in his task. Subjects were the agent, while the expert was the same trained experimenter. Indeed, the expert's role needs a learning process, while the agent only has to follow remote instructions. The task simulated a physical bimanual selection by touching simultaneously two different targets (cf Fig. 4), one with each hand. Since in mode (2) simple sketching provides no way to distinguish which hand the agent should use to select a given target, the expert used a color code to indicate which hand to use. No verbal communication in both modes was allowed. Each subject had to accomplish this task 20 times (10 times replayed from a pre-recording guiding phase performed by the trained experimenter, and 10 times with live manual guiding performed by the same trained experimenter) per mode. We stated hypotheses:

- H1: The subject's task completion from the completion of the guidance is faster in mode (1) ;

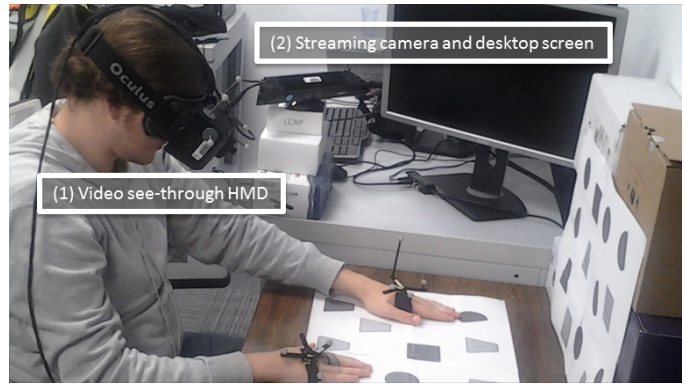
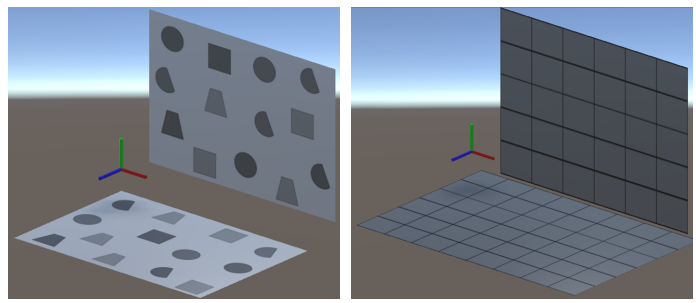


Figure 4: Illustration of a subject performing the pilot user study in mode (1) with the first targets' layout (L1).



(a) (b)

Figure 5: The virtual targets' spaces ((a): first layout (L1); (b): second layout (L2)), reproduced from the real agent's workspace, used in the expert's VE.

- H2: Mode (1) eases the mapping process between the guiding instructions and the physical task interaction space.

We ran this experiment twice with the same subjects' panel and conditions. First (L1) with a simple random targets' layout (cf. Fig. 5a), then (L2) with a more regular and dense targets' layout (cf. Fig. 5b) that seems more complex because of less visual and geometric cues.

5.1 Results

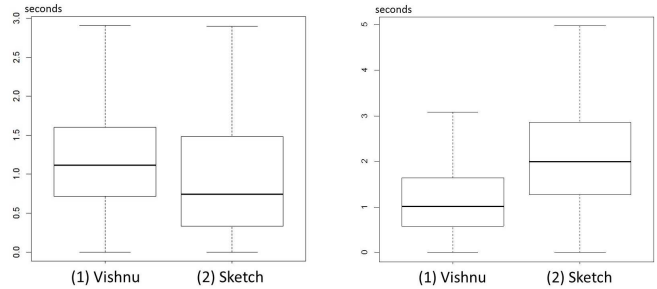
Results have been collected from 11 subjects aged from 23 to 40 ($mean = 30, sd = 5.7$). Participants have been selected according to their physiological acceptability of the video see-through HMD in order to minimize its effect in mode (1) with respect to mode (2).

For (L1), boxplots of figure 6a shows difference between the guidance completion and the subject's task completion. There is a significant difference between both modes: $F(1,432) = 3.93, p = 0.047$. Notice that we found no significant difference between the replayed iterations and the manual ones.

Concerning the qualitative results (cf Fig. 7a), mode (1) is preferred in terms of mapping easiness, originality as well as guide's presence. Mode (2) offers a better visual comfort.

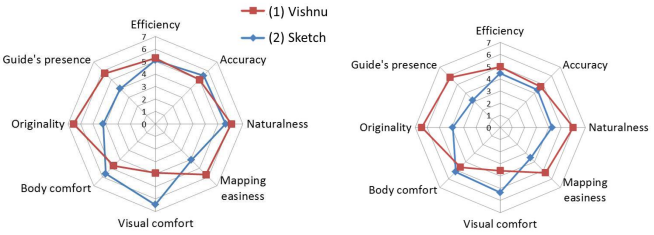
5.2 Discussion

We validated our hypothesis (H2). In particular, qualitative results are very encouraging concerning our technique (1), except for comfort (visual and global). This must be due to the use of the video see-through HMD that slightly alters the perception of the real environment and needs a learning process in order to match visual



(a) (b)

Figure 6: Boxplots of the difference between the guidance completion and the subject's task completion for the first (a) and the second (b) layout.



(a) (b)

Figure 7: Synthesis of qualitative results for the first (a) and the second (b) layout.

feedback and proprioception. It could also explain why we did not validate (H1) in a pretty simple task context (L1): there were only 13 targets by sides (26 in total, cf Fig. 5a).

Next section presents results in a more complex task (L2), i.e with a bigger set of more dense targets with uniform shapes. In this context, mode (1) could be significantly faster than (2) because the mapping complexity would override the perception issues generated by the use of the video see-through HMD in mode (1).

5.3 Second targets' layout

For (L2), we proposed targets amongst 78 uniform squares arranged in a regular grid (cf Fig. 5b). The idea is to remove spatial cues in the subject's real workspace in order to increase the mapping complexity of the correct targets in mode (2). It is similar to real world equipments such as computer racks or electrical panels that contain a lot of outlets arranged in a regular grid. Figure 6b shows task completion boxplots. In this more complex configuration, mode (1) is significantly faster than mode (2): $F(1,432) = 41.56, p < 10^{-9}$. Concerning qualitative results (cf Fig. 7b), they remain the same compared with the first layout (L1).

6 Conclusion

We presented our novel system that was developed to improve immersion and interactions for a remote expert helping an agent to perform a physical task. To achieve this, we proposed *Vishnu*, an interaction paradigm for remote collaborative maintenance based on two additional virtual arms coming out the agent's shoulders and controlled by the remote expert. This way, the expert can guide the agent in a natural way using arm and hand gestures and interact with virtual objects using his virtual arms as interaction guidance tools. Additionally, we proposed several other useful features that ease the use of the paradigm. The main advantages are the direct mapping of gestures for the agent due to collocation, and the interactive abilities of the remote expert to move virtual objects seen by the agent. A pilot user study validates our concept in terms of effectiveness and user preference: in a complex scenario, our technique is significantly faster compared with a desktop screen approach. Thus, future studies should evaluate its effectiveness in similar realistic scenario, such as the maintenance of electrical panels, or computer racks. Illustrative videos (*vishnu* prototype and pilot user study) are available online [1].

This work also opens perspectives regarding gestures-based guiding. It could be used in applications that explore ergonomic issues. This research field highlights the need to handle a correct re-targeting of virtual body parts according to an expert who controls them and an agent who sees them coming out of his own body. The approach described by Kulpa et al. in [9] seems appropriate to achieve this and should be adapted to our *Vishnu* paradigm in future work.

References

[1] Additional material online repository: `https://app.box.com/s/10q76uh8xgxc5juuxthuft0vg9hzett9`.

[2] M. Achibet, M. Marchal, F. Argelaguet, and A. Lécuyer. The Virtual Mitten: A novel interaction paradigm for visuo-haptic manipulation of objects using grip force. In *3DUI, 2014 IEEE*, pages 59–66.

[3] M. Billinghurst, H. Kato, and I. Poupyrev. The MagicBook: a transitional AR interface. *Computers & Graphics*, 25(5):745–753, 2001.

[4] M. Dou and H. Fuchs. Temporally enhanced 3d capture of room-sized dynamic scenes with commodity depth cameras. In *VR, 2014 iEEE*, pages 39–44.

[5] W. Huang, L. Alem, and J. Albasri. HandsInAir: a wearable system for remote collaboration. *arXiv preprint arXiv:1112.1742*, 2011.

[6] W. Huang, L. Alem, and F. Tecchia. HandsIn3d: augmenting the shared 3d visual space with unmediated hand gestures. In *SIGGRAPH Asia 2013 Emerging Technologies*, page 10. ACM, 2013.

[7] W. Huang, L. Alem, and F. Tecchia. HandsIn3d: supporting remote guidance with immersive virtual environments. In *HCI–INTERACT 2013*, pages 70–77. Springer, 2013.

[8] K. Kiyokawa, H. Takemura, and N. Yokoya. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. In *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings*, volume 6, pages 48–53. IEEE, 1999.

[9] R. Kulpa, F. Multon, and B. Arnaldi. Morphology-independent representation of motions for interactive human-like animation. In *Computer Graphics Forum*, volume 24, pages 343–351. Wiley Online Library, 2005.

[10] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi. The stretchable arms for collaborative remote guiding. In *ICAT-EGVE*, 2015.

[11] K. Robert, D. Zhu, W. Huang, L. Alem, and T. Gedeon. MobileHelper: remote guiding using smart mobile devices, hand gestures and augmented reality. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, page 39. ACM, 2013.

[12] F. Tecchia, L. Alem, and W. Huang. 3d helping hands: a gesture based MR system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on VRCAI*, pages 323–328. ACM, 2012.

[13] D. Tolani, A. Goswami, and N. I. Badler. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical models*, 62(5):353–388, 2000.

[14] J. Zillner, C. Rhemann, S. Izadi, and M. Haller. 3d-board: a whole-body remote collaborative whiteboard. In *Proceedings of the 27th annual ACM symposium on UIST*, pages 471–479. ACM, 2014.