HyPhIVE : A Hybrid Virtual-Physical Collaboration Environment

Senaka Buthpitiya and Ying Zhang
Carnegie Mellon University
Moffett Field, CA, USA
{*senaka.buthpitiya, joy.zhang*}*@sv.cmu.edu*

***Abstract*—Virtual world conferences have been shown to give users an increased sense of presence in a collaboration as opposed to teleconferences, video-conferences and web-conferences. Such telepresence encourages remote participants to engage in the collaboration. Current virtual world collaboration applications rely on mouse/keyboard interfaces to create pure-virtual collaborations. In this paper we propose HyPhIVE, a system to address hybrid collaboration between the physical world and virtual worlds. In hybrid collaboration scenarios, a group of people collaborate in the real world and others join them remotely via a virtual world. HyPhIVE uses non-intrusive mobile sensors to detect real world users' collaboration context such as their position, direction of gaze, gestures and voice. HyPhIVE projects the sensed real world collaboration into a virtual world in a way that collaboration patterns are preserved. Remote users join the collaboration using virtual world clients and interact with other users' avatars. User studies have shown that HyPhIVE effectively projects real world collaborations into a virtual world and it improves users' experience of remote collaboration.**

***Keywords*-collaborative virtual environments, ubiquitous computing, mobile computing.**

I. Introduction

Today's business, educational and research worlds exist in a global context, with people spread out across the globe participating in collaborations[1]. Providing an engaging platform for collaboration is essential for the success of the individual as well as of the group.

Over the last half-century various technologies have been developed as platforms for remote collaborations. These technologies include teleconference, video-conference and web-conference systems. Teleconference systems lack visual connection between remote participants which make the communication less personal and less interactive. Video conference systems project images captured from cameras onto screens. Though visual information is presented in a video conference, stationary camera positions cause loss of perspective for remote users. When the majority of users are in a single location and remote participants are by themselves, remote participants tend to watch the meeting rather than "participating" in the collaboration and feel disjoint from the group.

High-fidelity video conference systems such as HP's Halo system [2] and Cisco's Telepresence system [3] provide "life size" images of the remote participants in a conference, but these systems are plagued with many practical limitations. These systems are very expensive and require highly reliable and high-bandwidth networks for all participating sites. Furthermore, furniture, lighting, wall paint, etc. in all locations have to be meticulously arranged to convey the illusion of co-location.

Virtual worlds provide an alternative means for people to collaborate while being physically remote. Virtual worlds such as Second Life are typically graphical, immersive, interactive, computer-generated spaces that enable social networking and communications. A 3D virtual world provides participants with the feeling of telepresence and being part of a community of peers. This has a direct impact on the collaboration and encourages engagement especially for remote participants.

Current efforts in using virtual worlds for collaboration are purely "virtual". A participant use a virtual world client to control his/her avatar in order to collaborate with other participants. Dominant user interfaces for virtual world clients are still mouse and keyboard though alternative interfaces are available and have been tested [1].

In this paper, we propose a novel system HyPhIVE for hybrid collaboration between the virtual and real worlds. HyPhIVE stands for **Hy**brid **Ph**ysical and **V**irtual World **I**nteraction **E**nvironment and is pronounced as "High-Five". In hybrid collaborations, real world users proceed with their normal collaborations. HyPhIVE captures the essence of the real world collaboration using non-intrusive sensors and projects the collaboration into the virtual world. Remote users join the collaboration in the virtual world where they feel a sense of presence that motivates them to be more engaged. Main contributions of this paper are: 1) we present the first hybrid virtual/real collaboration system HyPhIVE; 2) we develop a model to quantify collaboration and 3) we describe an efficient algorithm to project collaborations from the real world to the virtual world.

The rest of the paper is organized as follows: Section II introduces the overall system architecture. Section III de-

[1]In this paper, we use *collaboration* to refer to interactions and communications between multiple people including meetings, lectures, seminars, presentations, discussions etc.

[2]http://h20338.www2.hp.com/enterprise/us/en/halo/index.html

[3]http://www.cisco.com/en/US/netsol/ns669/networking_solutions_solution_segment_home.html

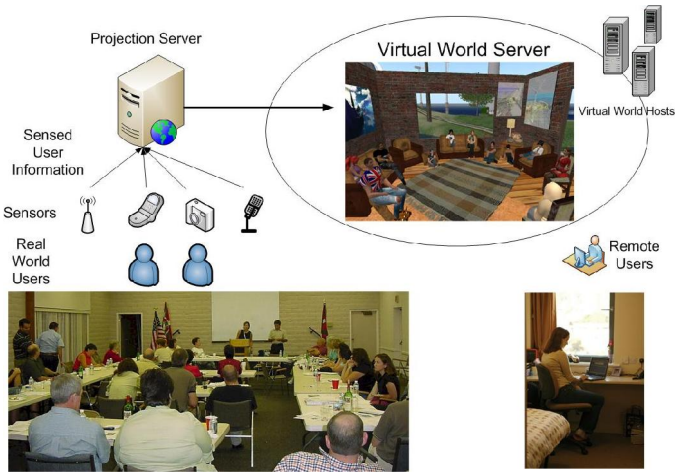

Figure 1. Architecture of the HyPhIVE system.

scribes how the system captures the real world collaboration through non-intrusive sensors. In Section IV we present the projection algorithm. We conclude the paper in Section VI with future work plans.

II. Hybrid Collaboration

Figure 1 illustrates the architecture of the HyPhIVE system. In the real world, each participating user carries his/her mobile phone during the collaboration. Sensors embedded in the phone, together with other sensor in the room, capture each user's gesture, position in the room, gazing direction and other features that are important for collaborations. This information is sent to a local *projection server* through the wireless network. The projection server combines the collected information and projects the real-world collaboration pattern into its corresponding virtual world representation. Based on the projected collaboration pattern, the projection server converts the user's movement and gestures into commands and sends commands to the virtual server so that avatars act accordingly.

Remote users use client software and join the collaboration in the virtual world. They move their avatars using standard keyboard and mouse interfaces and use microphone and speaker for voice communications. Activities of remote virtual users are shown to real world users by projecting the virtual meeting room onto screens in the real world meeting room. One alternative is to have all real world users wearing head-mounted display goggles and to augment users' vision with images of avatars from the virtual world. We can foresee that in the near future we could project virtual users' images into the real-world meeting room when 3D hologram technologies become more affordable and more mature. A hologram will make the real world users feel more natural and less awkward compared to the augmented reality solution of wearing goggles. We are also exploring the robotic display solution where a computer monitor mounts on a robotic platform. Faces of remote users are shown on the monitor and the monitor moves and turns while the remote users interact with real world users.

III. Sensing Real World Collaboration

Existing technologies either rely on body sensors such as magnetic tracking and exoskeleton tracking, or use 3D cameras to track body motion and project it into the virtual world. Placing body-sensors on participants in a meeting is not practical on a daily basis and using cameras to track users' motions is usually not reliable when multiple people are present. For hybrid real-virtual environments, we do not need very detailed body motion. Knowing users' positions in the room, their basic gestures (e.g. standing, sitting) and who is talking to whom is sufficient to create a reflection of the environment.

There is growing interest in the use of sensor-enabled "smart" mobile phones for people-centric sensing. Most smart phones have built-in sensors such as GPS receivers, accelerometers, WiFi receivers, embedded microphones and cameras. These sensors enable mobile applications to detect and infer users' context information such as their indoor location [2] and gesture [3]. In this work, we take advantage of the popularity of smart mobile phones to capture real world users' motion and activities that are important for collaboration.

We attach a bar code on the back of user's mobile phone. When entering a room, the bar code is scanned so that the projection server knows which user has just joined the collaboration and logs the user into the virtual world with his/her default avatar.

WiFi and Bluetooth receivers, together with accelerometers embedded inside mobile phones are used to detect a user's position in a physical room. We use the WASP algorithm [4] for room-level positioning. WASP is an extension to the Redpin algorithm [2] for congested Wi-Fi environments. WASP is based on fingerprinting of WiFi Radio Signal Strength (RSS) which gives a reasonable precision (about 85% to 90%) for room-level indoor locationing. Combined with accelerometer readings, we can locate a user's position inside a room once he/she "check-in" at the entrance.

We have developed a gesture detection algorithm based on accelerometer readings from mobile phones similar to [5], [6], [7]. Gestures that are interesting for collaborations include walking, sitting, standing, turning (left/right) and raising of a hand to get the attention of the person in charge of proceedings. In the current HyPhIVE system, we use only accelerometers embedded in users' mobile phones which allow the gesture detection to detect "walking", "sitting", "standing" and "turning" at an accuracy of approximately 90%.

HyPhIVE uses acoustic profiles collected from on-device microphones to decide collectively which users are engaged

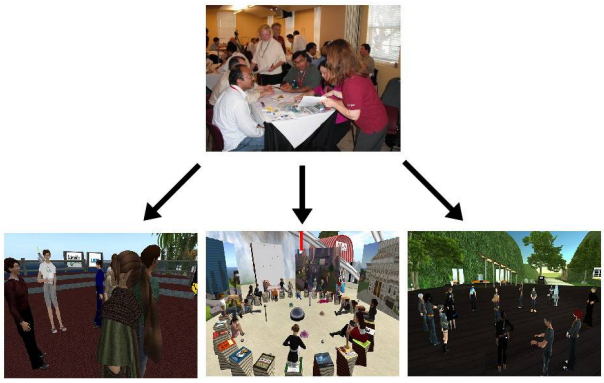

Figure 2. Projecting the real world collaboration into different virtual world settings.

in a conversation. This information will be used to help decide avatar orientation in the virtual world, i.e., real-world persons engaged in a conversation will be depicted by avatars facing each other in the virtual world. The information will also be used to decide on private and public conversations, allowing the system to make public conversations audible in the virtual world while the audio of private conversations will be suppressed.

IV. Collaboration Projection

In most cases, the real and virtual meeting rooms will have differing layouts. As illustrated in Figure 2, there are different ways to project a physical collaboration into a virtual world. One of many contributions of HyPhIVE is a collaboration model that quantifies "collaboration". With this model, we can compare whether two collaboration layouts are similar and search for a virtual collaboration layout that is most similar to the real world layout.

First, we model the Attention Index (I_{att}) between two people using the physical distance and direction of gaze. Next we can model the collaboration among a group of n people with an $n \times n$ matrix C, where $C_{i,j}$ is the amount of attention participant i pays to j.

A. Attention Index

Previous work in the field of Proxemics [8] identifies that the distance between two people is an important indicator of the level of attention between them. Psychologists have found that many users maintain a similar personal space around their avatars in the virtual world [9].

Extending this concept, we hypothesize that the angle at which people face each other is also an important indicator. [10] uses a two dimensional Gaussian function to model personal space based on four parameters: the distance between people, their face orientations, age, and gender. In our work, we consider only distance and face orientation to calculate the *attention index* (I_{att}) between two people,

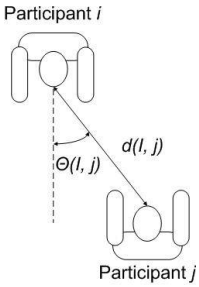

Figure 3. Attention Index based on the *distance* and *angle difference* between user i and j.

which is an abstract estimation of the attention being paid by one individual to another. We leave out age and gender from our model to simplify the calculations as a trade-off of accuracy for speed. We estimate I_{att} between two people i and j as:

$$I_{att}(i,j) = \frac{1}{\alpha \log d(i,j) + \beta\theta(i,j)}, \quad (1)$$

where $d(i,j)$ is the physical distance between user i and j, and $\theta(i,j)$ is the difference of their gazing angles (Figure 3). This model is chosen empirically following intuition. Other function forms may also capture the relationship between "attention" and distance/angle.

To fit parameters α and β in the I_{att} model, we gathered training data from a user study where human subjects were presented with multiple computer generated images of a virtual office environment populated with people. Subjects were instructed to "look directly into the center of the image" and rate "how much attention are you paying to the character wearing a red/blue striped tie that has a red spot on his forehead?" on a scale of 1 to 10. Figure 4 is one of the images shown to subjects during the evaluation. Thirty subjects participated in the study and evaluated 316 images. We displayed 15 randomly selected images to each subject and obtained 450 subjective evaluation scores. Since each user's rating is subjective, we fit the attention index function so that the *ranking* of images correlates with what the human subjects assigned rather than the actual scores of 1 to 10.

B. Modeling Collaboration Pattern

Attention Index quantifies the "collaboration" between two participants. For a group of n people, we use *Collaboration Matrix (C)* to quantify the collaboration among a group of people. Define the Collaboration Matrix C as a square matrix of $n \times n$, where

$$C_{i,j} = \begin{cases} I_{att}(i,j) & i \neq j \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The collaborative difference (or distance) between two layouts can be estimated using the collaboration matrices

Look directly into the center of the image.
How much attention are you paying to the character wearing a red/blue stripe tie and has a red spot on his forehead?
Rate on a scale of 1 to 10.

No attention	1	2	3	4	5	6	7	8	9	10	Full attention

Submit

Figure 4. Subjective evaluation of Interpersonal Interaction Model.

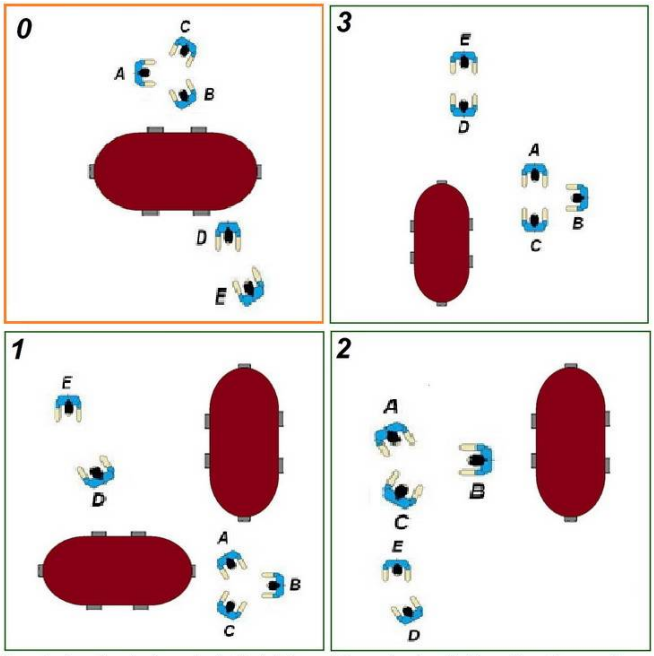

Figure 5. Subjective evaluation of similarities between different collaborations.

$C1$ and $C2$ of the two layouts. Denoting the collaborative distance as $D(C1, C2)$:

$$D(C1, C2) = \sum_{i=1}^{n} \sum_{j=1}^{n} |C1_{i,j} - C2_{i,j}| \quad (3)$$

The distance function is made as simple as possible to allow re-computation at real-time speeds for use in the mapping algorithm as explained in the following section.

A second user study was conducted to test the effectiveness of the attention index equation and the collaboration distance function in capturing the collaboration in a scene. The test data consists of 5 sets of manually created images. Each set contains 4 images where one image is considered as the original layout (image 0) and the other three are alternative layouts. Human subjects are asked to order the alternative layouts according to how they best mirror the collaboration occurring in the original layout (Figure 5). Results from approximately 40 participants were aggregated and averaged to rank the images in each set separately on collaborative similarities with the set's primary image. The image sets were sorted according to their similarity to the original image using the I_{att} equation and CD function. We calculate Spearman's rank correlation coefficient [11] to compare rankings of alternative images by human subjects and by automatic Collaboration Difference scores.

Spearman's rank correlation coefficient on this data set is ρ = 0.80. The probability for $\rho = 0.80$ in a data set of the magnitude used in the experiment is less than 0.01. In other words, the automatic method of measuring the similarity between two collaboration patterns correlates extremely well with human perception.

C. Mapping Algorithm

The mapping algorithm is the central component of the HyPhIVE system. HyPhIVE maps the real world collaboration into the virtual world. Given the collaboration matrix of a real world collaboration, the mapping algorithm generates a layout in the virtual world that is most similar to the real world collaboration. The virtual layout also needs to satisfy constrains of the virtual environment such as not placing avatars on top of each other and not placing avatars inside walls.

The number of possible avatar configurations within a virtual room is very large. As the number of real-world users (i.e., avatars to arrange) increases, the number of configurations increase exponentially. To allow the computation to occur in real-time, the algorithm takes a greedy approach in searching for the optimal avatar configuration. Figure 6 shows the mapping algorithm used in HyPhIVE. Denote the *configuration* of avatars in the virtual world as P where $P(i)$ includes i's coordinates and his/her direction of gaze. Starting with a randomly scattered avatar configuration, we modify the current configuration P by apply one action on one of the avatars such as move the avatar forward,

Input: Collaboration matrix of real world C_R, Current avatar arrangement P_0
Result: Avatars positions and orientations in VW
Push P_0 into a priority queue Q ;
STOP = false ;
while *STOP==false* **do**
  pop out P from Q;
  Calculate collaboration matrix C for P;
  foreach *avatar* i **do**
    foreach *movement possible for* i **do**
      Apply movement on i; P' = new avatar configuration;
      Calculate collaboration matrix C' for P';
      Calculate $D(C_R, C')$;
      Push P' into Q with $D(C_R, C')$ as its priority;
    end
  end
  retain the top B elements and flush Q ;
  if *top element's* $D(C_R, C') \geq D(C_R, C)$ **then**
    STOP = true ;
  end
end

Figure 6. Greedy mapping algorithm for avatar configuration.

backward, turn left or turn right. The top B new configurations (with the smallest "distance" from the real-world configuration) will be explored for further modification in the next iteration. To reduce the number of configurations after one modification, an avatar can only "hop" forward and backward and rotate by two degrees each time.

We performed a stress test to test the effectiveness and responsiveness of the mapping algorithm as the number of real world users increase. The test placed a certain number of participants in a static configuration and measured the time the algorithm would take to move the avatars into an acceptable and stable collaboration configuration. The acceptability of a configuration was decided by the collaboration distance between the two configurations. A stable configuration for this experiment was defined as a virtual world configuration which remains constant while the real world configuration remains unchanged. Each experiment is repeated 10 times with different randomly initialized configurations. Table I shows the average running time for configurations with different number of participants. As the number of participants increases, average search time increases. The *averaged per capita collaboration distance* value also indicates that the greedy search usually ends at a local minimum when the search space becomes too large.

V. Related Work

To the best of authors' knowledge, there is no existing work on hybrid virtual-physical collaboration environments as described in this paper.

RW Participants	*Avg. Execution Time (ms)*	*Avg. Per Capita CD*
5	0.14	4.40
10	2.66	10.41
15	13.64	17.06
20	44.19	25.71
25	75.35	36.60

Table I
Average time to reach a stable VW configuration from random initial configurations.

There has been much work into developing purely virtual collaboration environments such as the Open University project in Second Life[4]. These systems are collectively categorized as Collaborative Virtual Environments (CVEs). There has been extensive research on how to maintain consistency between users in distributed locations [12] and other system level aspects of CVEs. As HyPhIVE is designed to work over an existing CVE system it does not deal these issues directly.

Research into the social aspects of CVEs can be broadly categorized into two areas, 1) work that deals with the general aspects of avatar interaction and its effects on users [13]; 2) work that deals with improving/adapting a CVE for a particular tasks such as Educational Virtual Environments (EVEs) [14].

There has been some work into integrating augmented reality technology with CVEs [15], [16], but these approaches still face technological and economic limitations. From a technological stand point, the displays required for use in everyday life are not widely available. From an economic stand-point using augmented reality for collaboration does not seem viable in the near future as each participant would require a virtual reality display.

Previous work in social sciences, specifically in the field of Proxemics [8], has a direct bearing on this work . Hall shows that individuals tend to maintain a certain distance between each other when interacting, and that the level of interaction and intimacy can be deduced from the distance between the individuals. The converse of which can be considered as further evidence for the hypothesis on collaboration presented in this paper. Human controlled avatars in virtual environments have been shown to maintain interpersonal distances when interacting, mirroring their real world counterparts [9]. A mathematical model for interpreting personal space using distance, orientation, age and gender as parameters was developed by [17]. This work focuses on an individuals personal space and extends to interactions between two individuals yet it does not account for situations where multiple people are present. Previous work in automated arrangement of avatars in virtual worlds by [18] is limited to considering the distance between avatars

[4]http://secondlifegrid.net.s3.amazonaws.com/docs/Second_Life_Case_OpenU_EN.pdf

to gauge an optimal arrangement.

VI. Conclusion and Future Work

In this paper, we present HyPhIVE, the first hybrid collaboration environment between physical and virtual worlds. With non-intrusive mobile sensors, HyPhIVE captures real world users' collaborative contexts in the form of gestures, dialogues, directions of gaze and positions, and projects them into the virtual world. Remote users join the collaboration through virtual world clients. We develop a model for quantifying collaboration and an algorithm to project collaborations between environments. We show the effectiveness of the contributions by verifying collaboration projections with user feedback and by performance analysis. We plan to add additional sensors into the HyPhIVE system including cameras, microphones, eye gaze tracking cameras, etc. Additional sensors will allow us to capture extra gestures such as raising a hand to ask a question and to transmit relevant speech into the virtual world. As hybrid collaboration presents a new medium for interpersonal communication, we will also study how hybrid collaborations affect real world users and those joining virtually.

Acknowledgement

We acknowledge the efforts of Hsiuping Lin and Diwakar Goel for their work on initial iterations of this project. We also acknowledge the efforts of Patricia Collins, Heng-Tze Cheng, Feng-Tso Sun, Yi-Ting Yeh and the anonymous reviewers for their valuable feedback and suggestions on earlier drafts of this paper.

References

[1] J. Lee and I. Ha, "Real-time motion capture for a human body using accelerometers," *Robotica*, vol. 19, no. 06, pp. 601–610, 2001.

[2] P. Bolliger, "Redpin - adaptive, zero-configuration indoor localization through user collaboration," in *MELT '08: Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments*. New York, NY, USA: ACM, 2008, pp. 55–60.

[3] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," *American Association for Artificial Intelligence*, 2005. [Online]. Available: http://paul.rutgers.edu/~nravi/accelerometer.pdf

[4] H. Lin, Y. Zhang, I. Landa, and M. Griss, "Wasp: An enhanced indoor locationing algorithm for a congested wi-fi environment," in *Proceedings of The Second International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (MELT)*, Orlando, FL, Sep. 30 2009.

[5] F. G. Hofmann, P. Heyer, and G. Hommel, "Velocity profile based recognition of dynamic gestures with discrete hidden markov models," in *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*. London, UK: Springer-Verlag, 1998, pp. 81–95.

[6] M. Walter, A. Psarrou, and S. Gong, "Auto clustering for unsupervised learning of atomic gesture components using minimum description length," in *RATFG-RTS '01: Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*. Washington, DC, USA: IEEE Computer Society, 2001, p. 157.

[7] O. A. Baki, Y. Zhang, M. Griss, and T. Lin, "Mams: A mobile application to detect abnormal patterns of activity," in *Proceedings of the First Annual International Conference on Mobile Computing, Applications, and Services (MobiCASE)*, San Diego, California, Oct. 26-29 2009.

[8] E. T. Hall, "Proxemics," *Current Anthropology*, vol. 9, p. 83, Jun. 1968.

[9] D. Friedman, A. Steed, and M. Slater, "Spatial social behavior in second life," *Intelligent Virtual Agents*, vol. 4722, pp. 252–263, 2007.

[10] T. Amaoka, H. Laga, S. Saito, and M. Nakajima, "Personal space modeling for human-computer interaction." in *ICEC*, ser. Lecture Notes in Computer Science, S. Natkin and J. Dupire, Eds., vol. 5709. Springer, 2009, pp. 60–72.

[11] C. Spearman, "The proof and measurement of association between two things. by c. spearman, 1904." *The American journal of psychology*, vol. 100, no. 3-4, pp. 441–471, 1987. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/3322052

[12] C. Greenhalgh, J. Purbrick, and D. Snowdon, "Inside massive-3: Flexible support for data consistency and world structuring," in *CVE '00: Proceedings of the third international conference on Collaborative virtual environments*. New York, NY, USA: ACM, 2000, pp. 119–127.

[13] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Interpersonal distance in immersive virtual environments," *Personality and Social Psychology Bulletin*, vol. 29(7), pp. 819–833, Jul. 2003.

[14] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. I. Kramer, "Gestures over video streams to support remote collaboration on physical tasks," *Human-Computer Interaction*, vol. 19, pp. 273–309, Jul. 2004.

[15] M. Billinghurst and H. Kato, "Collaborative augmented reality," *Commun. ACM*, vol. 45, no. 7, pp. 64–70, 2002.

[16] A. C. E. for Service Providing in Cultural Heritage Sites, "Gestures over video streams to support remote collaboration on physical tasks," *Embedded and Ubiquitous Computing*, vol. 3207, pp. 273–285, Jul. 2004.

[17] T. Amaoka, H. Laga, S. Saito, and M. Nakajima, "Personal space modeling for human-computer interaction," in *Proceedings of The 8th International Conference on Entertainment Computing (ICEC)*. Springer Berlin / Heidelberg, September 2009, pp. 60–72.

[18] V. Quera, F. Beltran, A. Solanas, L. Salafranca, and S. Herrando, "A dynamic model for inter-agent distances," *From animals to animats*, vol. 6, pp. 304–313, 2000.