The ICOCOON Virtual Meeting Room: a Virtual Environment as a Support Tool for Multipoint Teleconference Systems

Aljosha Demeulemeester[1], Katriina Kilpi[2], Shirley A. Elprama[2], Sammy Lievens[3], Charles-Frederik Hollemeersch[1], An Jacobs[2], Peter Lambert[1], and Rik Van de Walle[1]

[1] Ghent University - IBBT,
Department of Electronics and Information Systems, Multimedia Lab,
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent, Belgium
[2] IBBT-SMIT, Vrije Universiteit Brussel, Pleinlaan 9, B-1050 Brussels, Belgium
[3] Alcatel-Lucent Bell Labs, Copernicuslaan 50, B-2018 Antwerpen,Belgium

Abstract. Globalization and increasing collaboration between remote teams drive the need for teleconference systems. However, currently no videoconferencing system matches the face-to-face experience for a business meeting with many participants in a flexible and affordable manner. In search for a better solution, we created a Virtual Meeting Room (VMR) application that visualizes key events detected using computer vision (e.g., participant entering the meeting room, talking, presenting) in a 3D virtual environment. The goal was to provide a good sense of overview to users when many meeting participants - represented by 3D avatars - from remote locations join a teleconference. In this paper, a technical overview of the working prototype - built using 3D game technology - is presented. Also, feedback from multiple user tests performed during the development of the prototype is discussed and presented as a set of recommendations. From the technical perspective, we found that existing 3D game technology is mature, affordable and contains the features needed to build the VMR application. From the users' and experts' feedback, we conclude that the VMR has merits as a teleconferencing support tool accompanying a video stream that conveys more detailed non-verbal communication of the active speaker.

1 Introduction

Teleconferencing systems are omnipresent in modern business environments. They range from basic phone conferencing systems, which allow many persons to join a phone conversation, to more elaborate setups using high-definition video streams displayed on large screens. Teleconferencing allows remote teams to collaborate without loss of time in traveling, which leads to cost and energy savings [1]. However, current-generation systems are lacking as users still report the need for face-to-face meetings, particularly when many participants are involved. Some of the central complaints about the current teleconferencing tools, on top of the complicated setups, are the lack of a clear overview of the entire meeting situation and the lack of the co-presence experience due to the way in which the audiovisual data is presented. The multidisciplinary ICOCOON project[4] aimed to tackle these issues by extracting knowledge by means of computer vision from audiovisual streams captured by a visual sensor network from multiple locations involved in a teleconference. This knowledge is then used to automatically select the most relevant video stream (e.g., someone presenting or asking a question) among all video streams from all locations to present to the user. Additionally, the project investigated the possibility of visualizing the captured knowledge in a virtual environment (VE)[5] resulting in our *Virtual Meeting Room (VMR)* application, which is the focus of this paper. Our approach aimed to visualize the high-level interpretation of captured knowledge in the context of a business meeting (see Section 3). The main goal of the VMR is to provide the same sense of overview (and more) one has during a conventional face-to-face meeting where all parties are present in the same location. The basic premise was that this could be valuable, especially in business meetings with many participants where only one participant is shown in a video feed to remote participants. Using a virtual environment, 3D humanoid avatars can convey actions of meeting participants in a way that can be interpreted intuitively by a user. Also, this approach offers a lot of flexibility on the way knowledge is visualized. This allows for an abstract representation of what is happening in the real world which facilitates the feeling of overview.

In this paper, we present a technical overview of our working teleconferencing prototype. Additionally, we present recommendations on the use of the VMR and the way knowledge should be visualized. The recommendations are derived from user evaluations and expert reviews, performed at different stages in the development of the prototype. The paper is structured as follows. After the overview of related work in Section 2, Section 3 discusses the semantics of captured knowledge and its representation in the VE. In Section 4, a high-level technical overview of the complete system is given. Section 5 describes the visualization module that renders and animates the virtual environment and the 3D avatars that represent the meeting participants. Section 6 describes the user research methods used and the set up of the proxy meeting technology. In Section 7, we formulate recommendations based on the feedback of the users and experts which we discuss further in Section 8.

2 Related Work

The next generation of teleconferencing goes beyond merely displaying a captured video stream of one or multiple meeting participants. The immersive 3D videoconferencing system of Kauff and Schreer [3] uses a shared virtual table environment. Participant video streams are modified using image-based rendering

[4] Immersive COmmunication by means of COmputer visiON [2]

[5] Bringing in physical elements into a predominantly virtual space is sometimes referred to as Augmented Virtuality.

to enable natural representation of gestures, eye contact and gaze. As a downside, the amount of participants is limited and they require a specialized setup for the purpose of teleconferencing.

Teleconferencing using a large amount of participants has been explored by IBM [1] using Second Life (SL), the massively multi-user online game-like 3D virtual world for social interaction. Users in SL are represented by a customizable 3D avatar in the virtual world and can interact with each other through spatial voice chat, text chat and avatar animation. The avatar is explicitly controlled by the user. The authors report positive feedback from the participants of a conference held using SL with over 200 participants. Kantonen et al. [4] enabled two types of participant: the first (virtual) participant uses a SL viewer and explicitly controls his avatar in the VE, the second uses augmented reality to visualize virtual participants in the physical meeting room. Hand tracking is performed to allow interaction with the augmented virtual world and to allow gesturing with their avatar. Regenbrecht et al. [5] present an augmented virtuality system for teleconferencing with video streams of participants being visualized inside the VE. Recent work by Lou et al. [6] focuses on informal communication in work settings. Their system called PresenceScape provides an immersive and interactive virtual office environment for the informal communication in the office space.

Using 3D avatars for communication has been extensively studied, both from the technical point of view as from the user perspective. A self-animated avatar in an augmented virtuality setup directly reflects the movement of a human in the real world. This allows using gestures and body language to convey a message [7]. The appearance (i.e., surface texture and geometry) of the avatar is usually predefined and the animation is done by manipulating the virtual bones of the character. A real-time motion capture system can track the position of all human limbs. However, to decrease the required network bandwidth, and to increase the amount of expressiveness in the avatar visualization, Arita et al. [8] symbolizes the raw motion capture data using knowledge on the communication context (e.g., teleconferencing, tele-teaching). The captured symbols, which represent intentions of communication, are sent over the network and transformed into high resolution motion data suitable to animate different avatars. This is comparable to our approach of detecting meeting participant actions that are relevant, sending this knowledge to all visualization clients and visualizing this in the most appropriate way in the VE. However, our system currently does not use detailed human motion capturing. This technology is not mature enough to capture human limb movement of many participants in one room full of obstructions (e.g., chairs, tables) without imposing many restrictions. Detailed non-verbal communication of participants is conveyed through the video stream that accompanies the VMR. As an alternative to motion capturing, the surface geometry of humans can be continuously captured using real-time 3D reconstruction from video streams [9]. These geometry sequences are implicitly animated and can be directly visualized in a VE. However, due to the lack of a control structure such as an animation skeleton, the geometry cannot easily be modified automatically to adapt the participant's virtual representation to the VE and other avatars (e.g., pointing or looking in the right direction). Also, 3D reconstruction requires more network bandwidth than streaming (symbolized) motion data.

While little research has focused on avatars in formal settings, Inkpen and Sedlin [10] investigated the use of 2D avatars for workplace videoconferencing. Their findings suggest that users are concerned how they are represented as avatars, but not about the representation of the ones they interact with. Focusing more on showing behavior than the design of avatars, Reidsma et al. [11] developed a virtual meeting room that can translate observations (such as gestures and gaze) from audio and video to animations in an avatar based virtual meeting room.

3 Meeting Observations in the VMR

A typical teleconference has participants from two or more locations. To enhance the feeling of co-presence, the avatars are represented in a single virtual (meeting) room which is the background for the visualization of captured knowledge. Similar to Reidsma et al. [11], we recognize that the observed knowledge from the physical world can be interpreted on progressively more complicated levels. For example, the exact position of the hand of a user can be detected. Depending on the position, this can be interpreted as hands being held up, which in itself can be interpreted as voting behavior or asking to speak. What we define as a *meeting observation* in our system are the high-level interpretations of the observed knowledge. These need to be visualized in the most recognizable way. Thus, the avatars in the VMR (see Figure 1) do not necessarily reflect participant actions directly. For example, the exact position of a meeting participant is not relevant so the result of the people tracking (see Section 4) does not need to be visualized exactly. Incidentally, it is not straightforward to map coordinates of multiple locations into one virtual location. However, proximity to certain objects (e.g., presenting board, other participant, etc.) is meaningful and is thus categorized as meeting observation. Using 3D humanoid avatars is an intuitive presentation form of meeting observations as opposed to text. This has the additional advantage that, when no high level interpretation of the collected knowledge can be made, the knowledge can be visualized directly to be interpreted by the user. For our earlier hand tracking example, this means that the avatar's hand position directly reflects the participant's hand position while leaving the user to decide on the meaning of the hand gesture.

In a teleconference scenario, knowing who is present is essential. This has to be interpreted as being part of the meeting in some way. Derived from presence are a number of meeting observations that we identified: participant enters the meeting, participant leaves the meeting temporarily (e.g., to restroom), participant exits the meeting permanently, participant re-enters after temporary leaving and participant joins unexpectedly. Additionally, it is visualized which participant is speaking, who is presenting and where each participant is looking

Fig. 1. The VMR application showing participant Erwin as presenting using a speech bubble with powerpoint logo; participant Victoria as speaking using a spotlight; participant Lucy with raised hand.

(e.g., is s/he paying attention). As a form of metadata about the physical world, detailed information of each participant (e.g., name, company, meeting role, etc.) is also shown in the VE.

4 Overview of the Teleconference System

The entire system can be subdivided in three main parts. First, information about each location is captured using a visual sensor network comprising of multiple spatially distributed cameras per location (smart rooms). This information is processed by a multitude of computer vision algorithms to attain knowledge about each location. The high-level vision algorithms each monitor a single property of one or multiple objects at a location (e.g., position of participants, are hands up/down, etc.) with high accuracy. The gathered knowledge is sent over the network to the central aggregator using JSON formatted messages. Second, the aggregator enriches and corrects the received knowledge and sends the consistent knowledge to the last subsystem, namely the visualization component (i.e., the VMR application). The latter consists of the visualization server and a visualization client per participant. The server is responsible for adapting the virtual world state to reflect what is happening in the physical world based on the received knowledge. Also, it distributes the virtual world

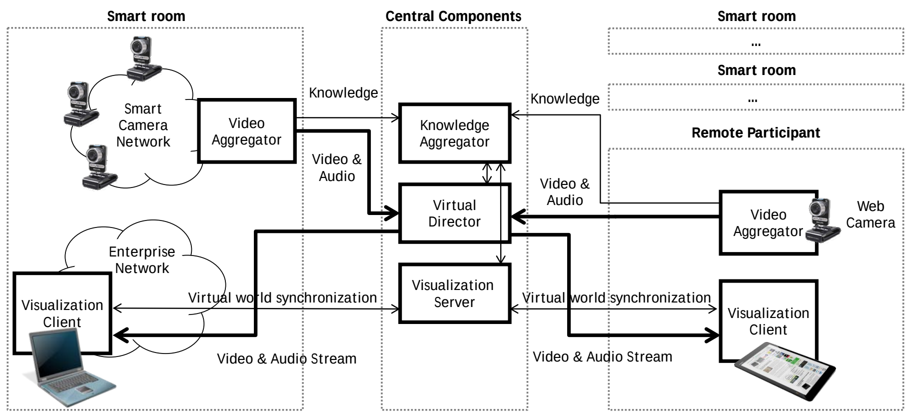

Fig. 2. High-level overview of the VMR teleconferencing system.

state to the visualization clients that each render their own unique view on the virtual world.

The architecture of the overall detection system consists of multiple smart rooms and the network infrastructure (see Figure 2). Each installed smart room comprises a calibrated smart camera network and one video aggregator. The smart cameras capture, resize, compress and send the video streams towards the video aggregator. The video aggregator extracts knowledge out of the captured video streams and sends this to the knowledge aggregator located in the network. In our smart room prototype, four overview cameras are mounted high and have a full view of the meeting room while three front-view cameras are mounted at eye height for capturing an optimal frontal view. Current extracted knowledge per location consists of a person's position and identification, gaze direction and activity status (e.g., sitting, walking, presenting, hands up/down). Four computer vision algorithms are used: face recognition by use of parabolic edge maps [12], people tracking [13], gaze detection [14] and hand detection and tracking. In our prototype, knowledge on people's position is used for extracting the behavior (e.g., entering, sitting, presenting) of the persons involved, based on the location of the person in the physical world (e.g., at the presenter board), the walking direction and speed and the size of each person. The people tracking module is also tightly coupled with the face recognition module that processes all people entering the meeting room and sitting at the table so that each moment in time the correct activity can be linked to the correct person. The person's gaze is extracted using a combination of face recognition and the upper-body detector module. In addition to the smart rooms, participants can also join from other locations. In this case, only the web cam of their personal computing device is used with a video aggregator running locally on that device (see Figure 2). The only difference for this kind of participant is that the types of meeting observation that can be made are reduced.

5 The VMR Visualization Client

In business meetings, it is common for participants to bring their own computing device (e.g., laptop, tablet, etc.). The VMR client is intended to run on these devices. This approach eliminates the need for dedicated hardware. Also, this allows participants to observe the meeting using the VMR even if they are not in one of the smart rooms. Therefore, the VMR client should run on the widest possible array of devices and require the minimum amount of memory and computing power possible. Because each client renders its own view on the virtual world, real-time 3D rendering, accelerated by common graphics hardware, is the preferred option to visualize the VE. As many avatars can be present in the VE, the avatar animation system should be highly optimized for speed. From these requirements, it was clear that modern 3D game middleware technology (i.e., game engine) is a good fit as it enables state-of-the art simulation and rendering while staying within the limits of commodity hardware. The Second Life system also offers many features relevant for our application but was not chosen because it is less flexible.

To manage and visualize the VE, the Unity 3D game engine[6] was selected as the foundation of the VMR application because of its range of features, its ease of use and its limited cost. The features critical to our application are a rendering engine for 3D graphics, scripting of object behavior and appearance, data driven animation and networking support. Equally important, Unity 3D deploys on most platforms, including iOS or Android mobile devices. The in-game Unity Editor enables the placement and configuration of objects in the virtual world. C# scripts are used to animate rigid objects. Unity 3D can load most popular mesh and animations formats and provides a system to play, blend, and cross-fade prerecorded (avatar) animation clips. No free open source engines were found that provide this combination of features. However, advanced path planning and Inverse Kinematics (IK) support is currently not present. The latter is used for automatically calculating joint rotations when directly animating hand and foot positions of a human avatar. Our VMR application requires only basic path planning and IK so this was not seen as a problem. Custom IK can be performed by overwriting the output of the Unity animation system for parts of the bone hierarchy.

To animate the VE and specifically the avatars, a set of controllers were designed that each modifies a (limited) set of virtual object properties. The navigation controller uses a graph based search algorithm to guide an avatar to an object or place (e.g., presenting board, chair) in the VE without collisions with other virtual objects. The animation controller manages a set of artist created animation clips. The clips are scheduled and cross-faded depending on the current desired end state (e.g., sitting, walking). The animation system needs at least one animation clip for each of the (currently six) states. All the avatars used have the same bone structure, avoiding the need for animation retargeting. Currently, avatar customization is limited to dynamically setting surface proper-

[6] Unity 3D: http://unity3d.com

ties (e.g., shirt color) or replacing the avatar's head. Render assets (i.e., geometry and textures, e.g., head scans) are loaded from a web server when needed. The procedural animation controllers use IK to directly control the head gaze and the avatar's hands. New controllers are easily added to the system. To ensure a consistent virtual world state, the visualization server calls the controllers running on the visualization clients. This uses less bandwidth than frequently sending the result of the controllers to the clients. Also, this allows for more customization of the visualizations at the client by the participant.

6 User Evaluation of the VMR

To improve and to evaluate our design of the next generation teleconferencing tool we used a multimethod approach. For this we introduced the Virtual Meeting Room to both users and experts with experience in teleconferencing. We conducted six group interviews with users after showing a video demonstration of the proposed system. Secondly, we conducted two expert reviews using heuristic evaluation with a working prototype of the system. Due to time limitations of the project, an evaluation of the prototype with users was not possible. This section describes our findings.

6.1 Video Demonstration and Group Interviews

The goal of the demonstration was to provide the users with the possibility to experience a close-to-realistic next generation teleconferencing system and create the feeling of immersion (see Figure 3). The demonstration was followed by group interviews. The setup of the meeting consisted of a table in a meeting room around which the users and two researchers (one leading the discussion, the other one observing) were sitting. The demonstration consisted of three screens: the first screen showing the video stream presenting the remote participants in the meeting (a video clip shot and edited beforehand), with next to it the VMR application demo film. In addition, one meeting participant (played by someone on the video stream) presented slides shown on a third screen. To create a connection between the participants in the VMR and the ones on the mocked up video stream, one of the researchers had a speaking role (i.e., asking a question). The videoconferencing system was reviewed by using variations of VMR visualizations as a starting point for discussion. The six group interviews had 3 - 5 participants each (mean group size = 3.8), which resulted in a total of 23 participants. The users' ages ranged between 22 and 50 years old (average = 33 years). The users came from different organizations and had different professional backgrounds such as a researcher, a manager and a historian. All users except for two had experience with both teleconferences and videoconferences. About 44% (n = 10) of the users had experience with virtual world applications.

Fig. 3. Left: Setup during group interviews with three screens: video stream, VMR and presentation slides. Right: Setup during expert reviews: the VMR, the video stream from remote locations and a feedback image of themselves.

6.2 Expert Reviews with Working Prototype

The prototype was reviewed by having a videoconferencing meeting between two locations (see Figure 3). The experts, each with a laptop, sat around a table. Each laptop displayed the VMR, a live video stream of the remote location and a feedback image of their own location. Both expert reviews were conducted with 3 - 4 people who all had a background in user experience, two of them especially in virtual world applications.

7 Results and Recommendations

After the group interviews on the mocked up technology, recommendations were formulated and applied to the prototype. The experts thus evaluated a system that had fundamentally been shaped by the user feedback from the group interviews. Similarly, after the expert reviews, the prototype was adjusted based on the findings of the experts. Although the mockup provided users with too much information, the expert reviews with the working prototype showed that this was no longer the case. This was most likely due to the reduction from three to two screens. Rather, the experts found the VMR to be a useful awareness tool during big meetings (i.e., more than 5 persons) where maintaining a good overview of the situation is harder. For instance, the VMR can provide support by showing what is going on in the real meeting room (e.g. who is speaking at the moment). Also, it gives a useful overview of who is present in the meeting, a feature, which was suspected to become more useful in bigger meetings with more remote participants. However, the findings suggest that the VMR might be of little use in a small videoconferencing meeting where a video stream covers all participants at once. Other changes after the user reviews include the implementation of a feedback image and the possibility to use a working interactive system (as opposed to the static mockup). In addition, the animations in the VMR that the users preferred were used during the expert reviews. In the next paragraphs, we list the recommendations pertaining to the VMR and highlight each recommendation with examples from our research.

VMR should provide an overview. Users and experts viewed the VMR application as an awareness tool that supports the users' knowledge on who is present in the room. Especially in a meeting with many locations and participants, users reported difficulties in keeping track of people's activities as they move in and out of the room. In addition, getting back on track after a moment of distraction was also viewed as being hard. Visualizations on presence status and meeting observations could support the user in having a more efficient meeting experience.

Design for immersion, though it is not always needed. The experts reported feeling immersed when using the prototype because of the interactivity between participants across rooms that the video stream facilitated. This had mostly to do with the video stream. However, depending on various aspects of the meeting such as relevance, commitment and interest, the users argued that immersion is not always needed. According to our users, it is common practice that when a user is not expected to actively contribute to a meeting, this passive follower might choose to multitask during the meeting. Although multitasking can be detrimental to some aspects of meetings, it is not always negative and can even positively impact meeting effectiveness and efficiency (cf. Lyons, Kim and Nevo [15]). The VMR can help the multitasking participant to get back on track.

Provide background information about users. In teleconferencing meetings, especially with unknown participants, people sometimes miss someone's name, job title or other relevant information. Stemming from this need, the users suggested displaying basic useful information (i.e., name, job title and company) about other meeting participants. We chose to display a person's name close to each avatar, and allow the retrieval of additional information in a pop up when hovering over a specific avatar with the mouse cursor (see Figure 4).

Visualization should be instantly clear and leave no room for misinterpretation. To allow the participant to focus on the meeting itself, all functionalities of the VMR should be instantly clear. Although some users found the size of the pie chart (see Figure 4) intimidating, most users appreciated the pie chart because it provided an overview of the location each participant belongs to. However, the functionality of the prototype, in which one could request additional information about avatars was not clear enough for the experts. Therefore, we suggest using visual cues (for instance a button) to make the functionality more explicit.

Visualizations should not distract. The users found the moving and the blinking of the animations distracting. Some users however pointed out that in certain situations a blinking animation could help to draw the attention of the participant who is otherwise preoccupied. To achieve a balance, we decided to show animations at certain events, such as the event of entering by showing a walking avatar. Some users considered this animation interesting to watch when participants were waiting for the meeting to start. However, when a participant joined a meeting at a later point in time, this animation had to be subtle and non-distracting, which was visualized by the avatar instantly appearing into an empty chair.

Allow users a degree of control. The users indicated that they should be given more control over the information presented in the VMR. For instance, some users argued that some information should be made optional, since it is not always needed. By giving users the possibility to customize which types of information they want to view, their feeling of control is increased. Users should also be allowed to control the appearance of the avatar representing them. While some users reported being satisfied with the minimum adjustments of the default avatar (i.e., avatar's shirt color matches shirt of meeting participant), others wanted more control over the avatar's looks to feel comfortable with their representations. Therefore, allowing customization of avatars based on user needs is recommended.

Connect with reality. To help users understand the VMR and to enhance immersion, observations visualized in the VMR should correspond to reality. This means that any gesturing or actions tied to a direction should translate realistically to the VMR. For example, when a person A is facing person B, the two avatars should also face each other in a similar way. In addition, the seating arrangements of the avatars should correspond to the same sitting order of the participants in same physical locations. During the group interviews, the importance of this connection was not a central topic to be investigated. At the expert reviews, however, this lack of connection turned out to be of importance, when dealing with own feedback image, on the VMR and on the video stream.

8 Discussion

The VMR was primarily viewed by both users and experts as a possible support tool for bigger meetings with multiple locations and more than six participants. For instance, the VMR can be used to convey information about any meeting participant, but also to provide an overview of the other locations. In addition, the VMR can offer a viable solution in situations where video stream is not possible to certain parties of the meeting (e.g. due to low bandwidth).

We have contributed by exploring the kinds of support people would benefit from in meetings held in VMRs, as proposed by Hasler, Buecheler and Pfeifer

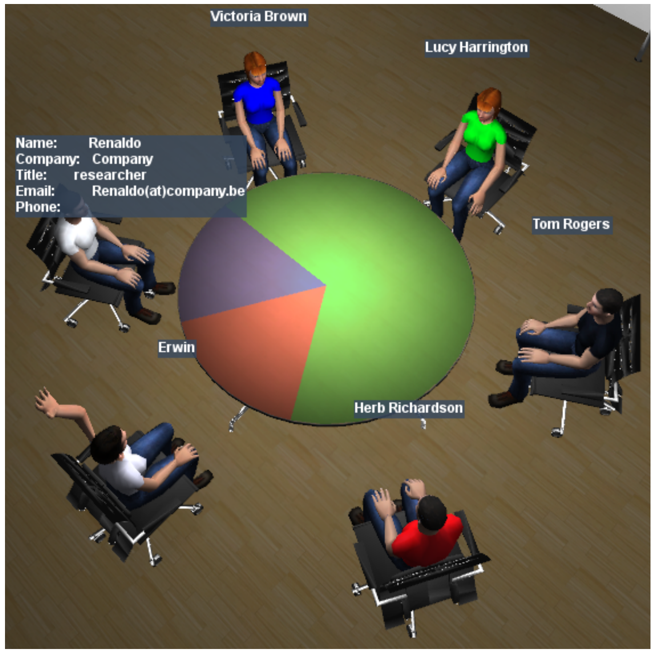

Fig. 4. Top down view in the VMR showing the pie chart table overlay that indicates the location of each participant. Also, the user information of Renaldo is shown.

[16]. Although experts reported feeling immersed, they, like the users, found that full immersion is not always needed (cf. Bowman and McMahan [17]), depending on the type of meeting. The design of a VMR should be clear and not contain elements that are distracting to the meeting participants. Also, the information presented in the VMR should be coherent with reality. Finally, we recommend to allow users more control, both regarding the amount of information presented and the customization of avatars.

Moreover, we recommend using a proxy technology tool to allow users the possibility to simulate the future technology. Although only a simulation, the proxy works as a starting point for discussion. It also allows the user researchers to contribute to the development of the technology with user insights, while the developers can continue working on the VMR. Through the iterative process, the prototype could be adapted incrementally, leaving enough time for the changes to be made and allowing the implemented recommendations to be investigated.

Because the prototype of the VMR used for the expert reviews only allowed testing with small meetings (e.g. up to five people), our findings are limited to smaller meetings. Future research could include testing the prototype with a larger group of people in more than one remote locations. This would allow us to investigate whether the usefulness of the VMR increases with more people as we hypothesize. Also, since meeting dynamics might differ in large meetings, the amount or type of information needed could change. For instance, we could investigate the added value of providing additional information, such as the agenda points still to be covered.

Another important aspect of the ICOCOON system is immersion. Even though some research participants reported that it is not always necessary to be immersed during meetings, it would be interesting to study how different camera views and perspectives, specifically first person view, would affect the level of immersion. Further research should be conducted to investigate these research questions.

9 Conclusion

Game technology provides a good foundation for an experimental VMR application. Our choice of game engine was a good fit for prototyping, incremental design and implementation. Unity 3D strikes a good balance between ease of use and the amount and the complexity of the provided features, which results in a limited amount of software development knowledge needed. From our discussions with users and experts, we conclude that our VMR approach has merits as a teleconferencing support tool without the need for an immersive experience. We strongly feel that the benefit of the VMR rises with increasing amount of participants and locations. However, we could not confirm this due to the limited amount of test users in one session. Essential to our application are avatars that are immediately recognizable as their corresponding participant. Our future research focuses on pre-meeting scanned, photo-realistic avatars. In our current prototype, most non-verbal communication is conveyed through the video stream because our users stressed the fact the redundant information is confusing. Additional research could identify how much of the physical world needs to be brought into the virtual world to make the video redundant. This would change the VMR from a support tool to a communication tool.

Acknowledgements

We would like to thank the following people for their contribution: Erwin Six, Kris Naessens, Donny Tytgat, Vinay Namboodiri, Maarten Aerts, Lizzy Bleumers, Koen Willaert, Bram Lievens, Ralf De Wolf and Chris Poppe. Also, we thank the interviewed participants. The research activities in this paper have been performed in the context of the ICOCOON project[2] and were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientic Research-Flanders (FWOFlanders), the Belgian Federal Science Policy Office (BFSPO), and the European Union.

References

1. Linden Lab: How Meeting In Second Life Transformed IBM's Technolgy Elite Into Virtual World Believers. Technical report (2009)

2. ICOCOON: http://www.ibbt.be/en/projects/overview-projects/p/detail/icocoon-2. IBBT. (2010)
3. Kauff, P., Schreer, O.: An immersive 3D video-conferencing system using shared virtual team user environments. In: Proceedings of the 4th international conference on Collaborative virtual environments. CVE '02, New York, NY, USA, ACM (2002) 105–112
4. Kantonen, T., Woodward, C., Katz, N.: Mixed reality in virtual world teleconferencing. In: Virtual Reality Conference (VR), 2010 IEEE. (march 2010) 179 –182
5. Regenbrecht, H., Ott, C., Wagner, M., Lum, T., Kohler, P., Wilke, W., Mueller, E.: An augmented virtuality approach to 3D videoconferencing. In: Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on. (oct. 2003) 290 – 291
6. Lou, Z., Bouwen, J., Willaert, K., Van Broeck, S., Van den Broeck, M., Zubic, S., Van Raemdonck, W., Van Herreweghe, H., Dams, D.: PresenceScape: Virtual World Mediated Rich Communication. Bell Labs Technical Journal **16**(4) (2012) 219–242
7. Dodds, T.J., Mohler, B.J., Blthoff, H.H.: Talk to the Virtual Hands: Self-Animated Avatars Improve Communication in Head-Mounted Display Virtual Environments. PLoS ONE **6**(10) (10 2011) e25759
8. Arita, D., Taniguchi, R.: Real-time Human Proxy: An Avatar-based Communication System. J. UCS **13**(2) (2007) 161–176
9. Lien, J.M., Bajcsy, R.: Skeleton-Based Compression of 3-D Tele-Immersion Data. In: Distributed Smart Cameras, 2007. ICDSC '07. First ACM/IEEE International Conference on. (sept. 2007) 347 –354
10. Inkpen, K.M., Sedlins, M.: Me and my avatar: exploring users' comfort with avatars for workplace communication. In: Proceedings of the ACM 2011 conference on Computer supported cooperative work. CSCW '11, New York, NY, USA, ACM (2011) 383–386
11. Reidsma, D., op den Akker, R., Rienks, R., Poppe, R., Nijholt, A., Heylen, D., Zwiers, J.: Virtual meeting rooms: from observation to simulation. AI Soc. **22**(2) (October 2007) 133–144
12. Deboeverie, F., Veelaert, P., Philips, W.: Face analysis using curve edge maps. In Maino, G., Foresti, G.L., eds.: Lecture Notes in Computer Science. Volume 6979., Springer (2011) 109–118
13. Grünwedel, S., Jelaca, V., Niño Castañeda, J., Van Hese, P., Van Cauwelaert, D., Veelaert, P., Philips, W.: Decentralized tracking of humans using a camera network. In Roning, J., Casasent, D., eds.: PROCEEDINGS OF SPIE, THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING. Volume 8301., SPIE (2012) 9
14. Marin-Jimenez, M., Zisserman, A., Ferrari, V.: heres looking at you, kid. detecting people looking at each other in videos. In: British Machine Vision Conference. (2011)
15. Lyons, K., Kim, H., Nevo, S.: Paying Attention in Meetings: Multitasking in Virtual Worlds. In: First Symposium on the Personal Web, Co-located with CASCON. Number 2005 (2010) 7
16. Hasler, B.S., Buecheler, T., Pfeifer, R.: Collaborative Work in 3D Virtual Environments : A Research Agenda and Operational Framework. In: HCI. Volume 12. (2009) 23–32
17. Bowman, D.A., McMahan, R.P.: Virtual Reality: How Much Immersion Is Enough? Computer **40**(7) (July 2007) 36–43