roomXT: Advanced Video Communication for Joint Dining over a Distance

Felix Heidrich[1], Kai Kasugai[1], Carsten Röcker[1], Peter Russell[2], Martina Ziefle[1]
[1]Communication Science, Human Technology Centre (HumTec), RWTH Aachen University
[2]Chair for CAAD, Architecture Department, RWTH Aachen University

{heidrich, kasugai, roecker, ziefle}@humtec.rwth-aachen.de
russell@caad.arch.rwth-aachen.de

***Abstract*—This paper illustrates the development of the *roomXT* system, a mixed reality communication system for the home domain. *RoomXT* virtually extends a user's physical environment by providing a "life-like" communication channel for informal and spontaneous interactions. We use a wall-sized display together with head-tracking and 3D rendering to achieve visual contiguity between the real and a virtual environment. As one possible application of this system, we set up a remote dining situation for two users. The dinner table is visually extended into the virtual environment in which a live video stream of the dining partner is shown. Great care was taken in regard to the aesthetic and device-free integration into the living room. First impressions of potential users support the validity of our design decisions.**

Fig. 1. *roomXT* video chat settings.

I. Introduction

The way we work and live changed dramatically over the last decades. These profound social changes also affected the composition of families, which becomes especially evident in the growing number of elderly persons living alone. Already today, a high number of older or chronically ill people live on their own, without support by their families. Over the last decades, there was a considerable increase in the number of single households, especially in the group of elderly people, and this trend is expected to continue in the coming years. Recent statistical data shows, that the percentage of elderly people living alone increases dramatically with age. In this context, pervasive healthcare systems are a promising solution for providing individualized medical support to a growing number of elderly and disabled people living alone. By enhancing physical spaces with information, communication and sensing technology and making them sensitive and responsive to the presence of people, pervasive healthcare systems provide assistive services in different areas of life (see, e.g., [1] or [2]).

II. Technology-Mediated Social Interaction

While pervasive healthcare technologies enable elderly and disabled people an independent life in their own home, they also bear the risk of reducing social contacts to a bare minimum. A number of projects have addressed this problem by designing special communication devices for supporting awareness and informal communication among distributed family members and friends. In comparison to traditional communication technologies, which were designed to support goal-oriented information exchange, these new systems focus on supporting social interactions between users in order to foster a feeling of connectedness within a group of closely related peers. While most early systems, like, e.g., TeamSCOPE [3], Sideshow [4], or Team Portal [5] were designed for supporting awareness and informal communication in distributed working groups, a variety of more recent approaches have a stronger focus on connecting users in technology-enhanced home environments. A well-known example is the FamilyPlanter [6], an ambient awareness system especially developed to be used by family members living apart. The system uses infrared and ultrasonic sensors for monitoring the presence of users and automatically transmits individual user states to selected family members. At the remote side, the transmitted information is visualized in the user's local environment via an ambient display. A similar approach was taken by Hindus et al. [7] with the Lampshade Intentional Presence Lamp (IPL), a decorative artifact that serves both as a lamp and a communication device. In contrast to the FamilyPlanter, the presence of people is not automatically captured. Instead, presence information is only communicated when users explicitly activate their Lampshade device. Other systems for supporting awareness in small intimate groups include Digital Family Portraits [8], Gleams of People [9], and Faint-Pop [10].

Today, the majority of awareness systems for home usage are built around dedicated communication devices (e.g., [11], [12]), which are either stationary systems available in the users' environment or mobile artifacts for personal usage.

PervasiveHealth 2012, May 21-24, San Diego, United States

DOI 10.4108/icst.pervasivehealth.2012.248679

While this appears to be a promising design choice, empirical evidence suggests that many users prefer calm and unobtrusive technology that blends into the existing environment instead of having visible communication devices within their living spaces [13], [14]. Hence, we aimed at designing a shared communication space by virtually extending the physical environment and thereby providing users with a "natural" communication channel for informal and spontaneous interactions. The following section illustrates how the overall concept was materialized in a prototype system called *roomXT*.

III. roomXT

In [15] we describe the vision of spatially extending rooms by using a large screen displaying 3D scenes with a perspective that matches the viewer's head position. In the following, we will give more details on this concept as it serves as the starting point for our approach.

The effect of virtually extending a room is similar to trompe l'oeile paintings [16], which create the optical illusion of three-dimensional objects or scenes on a canvas. Often, these paintings require to be viewed from one specific position in order to make the illusion appear realistic. With our *roomXT* system, we overcome this limitation by tracking the viewer's head position and continuously redrawing the perspective of a computer-generated scene. When a user moves her head, the perspective adjusts accordingly. This does not only allow viewing the scene in different angles, but also adds to the illusion of space [17].

Recent technical developments in the display market make concepts like *roomXT* increasingly feasible, even for home usage. We believe that within the next one or two decades, large interactive screens similar to the one shown in Figure 2 will be an integral part of many domestic spaces. Once displays reach the dimensions of room elements, like walls or at least windows, they do not only become part of the surrounding, but can also replicate the physical environment itself.

RoomXT can change the perception of the room in various ways. The physical space is linked to the virtual space by using a common vanishing point and by matching spatial dimensions, thus creating visual context, which leads to the impression of a connected room. It furthermore changes the room's properties by adding different functions to it. What was once a living room now becomes a shared space, which can, for example, act as a doctor's waiting or examination room (e.g., for first visual checkups and interviews), or any other room, in which social interaction takes place.

One natural limitation of *roomXT* is that, using standard display technology, the perspective can only be correct for one user at a time. Figure 2 shows an illustration of the desired impression: Two users stand in front of a large-scale display that shows two remote communication partners in a joint virtual space. The perspective matches the camera that took the photo of the scene (i.e., it is correct for the location of a third user), while both users in the photo would only see a distorted perspective. This limitation might be overcome with future display technology[1] [18], however it is of no relevance in our scenario, which focuses on single households.

On the technical level, various other video communication system like MAJIC [19] and BISi [20] were developed, that offer very high quality and a sophisticated system setup. From a purely technical point of view, such systems might be more sophisticated than the current *roomXT* prototype. Compared to many other approaches, our focus was not only on technical aspects, but in particular on a aesthetically pleasing and unobtrusive integration of a mixed reality conferencing system into the domestic environment. This should allow a more realistic evaluation of the system, especially in terms of user acceptance. As the wall-sized display becomes a part of the room, our system should enable direct and device-less interaction between remote communication partners. We are not aware of any other system that was designed with these goals in mind.

For our prototype we chose the dinner scenario for various reasons. Consumption of food and social interaction is strongly connected. By using a common and well-known situation, we want users to overcome potential initial restraints of using the system.

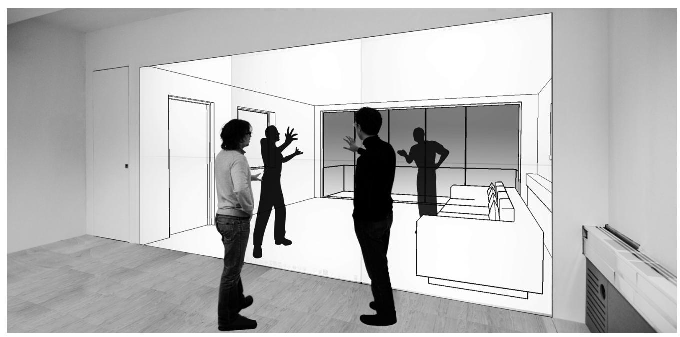

Fig. 2. Illustration of the *roomXT* concept.

A. System Setup

The system consists of two sites; both are located in the same building. The first site, the living lab (Figure 1), offers a rich video chat. The user experience at the remote site (Figure 3) is limited in some respect. Regarding experimental flexibility and also financial feasibility, we did not implement the *roomXT* setup at both sites.

1) Living Lab: The living lab contains a wall-sized display (4.8 x 2.4m, 3072 x 1536 pixels), an area of which is used to display the incoming video stream. The display wall is implemented using rear projection, so that the room resembles a living room rather than a media space. A table (90 x 90cm) is placed in front of the center of the display wall.

The user's head is tracked using a Microsoft Kinect[2] sensor and the corresponding Microsoft Kinect SDK. Performing a

[1] i.e., "200-inch glasses-free 3D display" ("http://www2.nict.go.jp/pub/whatsnew/press/h22/110125/110125_e.html")

[2] http://www.xbox.com/kinect

smoothing on the skeletal data gives us a jitter-free head position. In our setup, the Kinect is located behind the user at a distance of about 3 meters. In contrast to marker-based tracking systems, this marker-free approach does not disturb the appearance of the living room.

For audio playback we used two small consumer level speakers, located on the floor on both sides of the table. Audio and video were captured using a consumer level wide-angle web-camera with an integrated microphone.

2) Remote Site: While we do not focus on the user experience at the remote site, the quality of the video communication setup is high enough to enable normal human conversation without problems or discontinuities caused by technical limitations.

The remote site is located in a common office room (Figure 3). We placed a consumer level wide-angle web-camera directly above a 24-inch display located at one end of a table. In this setup, we used a longer table (168 x 90cm) that was only partly captured by the camera. The camera's optical axis is parallel to the table at a height of 115,5cm, which is approximately the height of a user's eyes when sitting at either of the two tables. We found, that the described configuration is a reasonable tradeoff between display size and offset between the dining partner's eye location in the video stream and the camera. This offset is small enough to give users in the living lab the feeling of being directly looked at. Carefully adjusting the camera position is necessary, because the final video stream should match the actual view of the dining partner as close as possible.

For audio recording we use a large-membrane microphone, located close to the user without being captured by the camera. Audio playback was achieved using the same consumer level speakers as in the living lab.

We implemented a green-screen behind the user at the remote site, which allowed us to perform simple background segmentation in the captured video stream. Additionally, a green tablecloth covered the table, so that hands, plate, and drinks on the table could be masked in the video stream of some conditions. For future applications we think that a consumer depth camera like a Microsoft Kinect could be used to easily mask the captured images. In this case, a cumbersome green-screen setup would not be necessary and users at both ends could have a more similar experience.

3) Software Infrastructure: We developed a Java software that is used for video transmission and rendering at both sites. Video images with a resolution of 640 x 480 pixels are captured from a USB camera and compressed as JPEG images using the turbo-jpeg library at 30 frames per second. This motion JPEG approach is easy to implement and enables a very low latency video transmission (30-50ms approximate delay). Images are transmitted via Ethernet to the remote system where they are decompressed and stored as OpenGL textures.

In the case of the living lab, the software then performs simple background segmentation on the textures. We also compute a projection matrix that is based on the user's head position. This allows us to render a virtual 3D table (90 x 90cm, like the real counterpart) that is shown perspectively correct for the user (Figure 1, right). The segmented video texture is then rendered at the far end of the virtual table. Finally, we align the edges of the captured table (i.e., the one with the green table cloth) and the virtual table. This gives the impression that the remote dining partner actually sits on the virtual table.

Our software supports rendering 3D scenes in the background, which also react to the user's head position. We implemented a virtual forrest scene that adds more spacial context. This further enhances the feeling of an extended room (Figure 1, left). When a user moves, trees that are close to her will move faster than more distant trees. In the future, other backgrounds could be implemented as well.

Audio transmission is accomplished with the Apple iChat application. Using Apple iChat in our local area network results in good audio quality with acceptable acoustic echo cancellation. Due to the jitter-free throughput in our local wired network, a time synchronized video playback or synchronization of audio and video during playback is not required. Simple decoding and displaying of the video images together with audio from iChat is sufficient for a flawless videoconference experience. The quality of the video conferencing system seems to be superior to many consumer level solutions. Especially the low latency is essential for natural human communication. However, the camera resolution and the pixel density of the wall-sized display are somewhat limited.

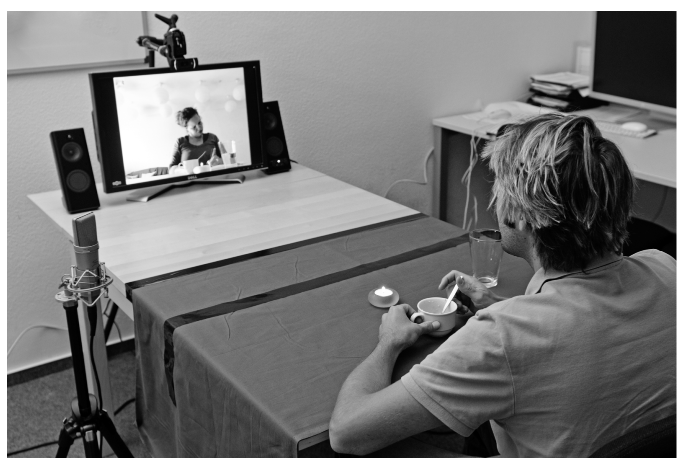

Fig. 3. Setup at the remote site.

IV. Conclusion

In this paper, we presented a video-based communication system for the home domain. It extends a user's physical environment by providing a natural communication channel for social interactions. Our prototype allows us to better understand how the system is perceived by potential users.

Several user studies are currently being conducted. First results of an explorative evaluation with a small group of users indicate the relevance of our design goals of an aesthetic and well integrated system.

V. Limitations and Future Work

In [21] a free viewpoint video system is presented. It uses multiple (depth-)cameras to capture a 3D mesh of a person in realtime. Such a system could be used to overcome current limitations. By rotating the mesh, gaze directions could be exactly preserved. Furthermore, it would allow the 3D mesh to be perspectively projected and thus give a more realistic 3D view of the communication partner.

In addition, other means of asynchronous communication (e.g., stored messages, drawings, or videos) could also be added, for example as developed in [11] for dedicated devices. Finally, we will complete our comprehensive user studies to evaluate our prototype and identify critical aspects.

Acknowledgments

We would like to thank all dinner volunteers. Many thanks also to Dominik Schlütter for research assistance. This work was supported by the excellence initiative of the German federal and state government.

References

[1] C. Röcker and M. Ziefle, Eds., *Smart Healthcare Applications and Services: Developments and Practices*. Niagara Falls, NY: IGI Publishing, 2011.

[2] C. Röcker and M. Ziefle, Eds., *E-Health, Assistive Technologies and Applications for Assisted Living: Challenges and Solutions*. Niagara Falls, NY: IGI Publishing, 2011.

[3] C. Steinfield, C. Jang, and B. Pfaff, "Supporting virtual team collaboration: The TeamSCOPE system," in *Proceedings of GROUP '99*. New York, NY, USA: ACM, 1999, pp. 81–90. [Online]. Available: http://doi.acm.org/10.1145/320297.320306

[4] J. J. Cadiz, G. Venolia, G. Jancke, and A. Gupta, "Designing and deploying an information awareness interface," in *Proceedings of CSCW '02*. New York, NY, USA: ACM, 2002, pp. 314–323. [Online]. Available: http://doi.acm.org/10.1145/587078.587122

[5] D. Atkins, D. Boyer, M. Handel, J. Herbsleb, A. Mockus, and G. Wills, "Achieving speed in globally distributed project work." in *Human Computer Interaction Consortium*, Frazer, CO, USA, 2001.

[6] K. Kuwabara, T. Ohguro, T. Watanabe, Y. Itoh, and Y. Maeda, "Connectedness oriented communication: Fostering a sense of connectedness to augment social relationships," in *Proceedings of SAINT '02*. Washington, DC, USA: IEEE Computer Society, 2002, pp. 186–193. [Online]. Available: http://dl.acm.org/citation.cfm?id=646465.691407

[7] D. Hindus, S. D. Mainwaring, N. Leduc, A. E. Hagström, and O. Bayley, "Casablanca: designing social communication devices for the home," in *Proceedings of CHI '01*. New York, NY, USA: ACM, 2001, pp. 325–332. [Online]. Available: http://doi.acm.org/10.1145/365024.383749

[8] E. D. Mynatt, J. Rowan, S. Craighill, and A. Jacobs, "Digital family portraits: supporting peace of mind for extended family members," in *Proceedings of CHI '01*. New York, NY, USA: ACM, 2001, pp. 333–340. [Online]. Available: http://doi.acm.org/10.1145/365024.365126

[9] T. Ohguro, S. Yoshida, and K. Kuwabara, "Gleams of people: Monitoring the presence of people with multi-agent architecture." in *Proceedings of PRIMA '99*, ser. LNCS, H. Nakashima and C. Zhang, Eds., vol. 1733. Springer, 1999, pp. 170–181. [Online]. Available: http://dblp.uni-trier.de/db/conf/pricai/prima1999.html#OhguroYK99

[10] T. Ohguro, K. Kuwabara, T. Owada, and Y. Shirai, "FaintPop: in touch with the social relationships," in *New Frontiers in Artificial Intelligence*, ser. LNCS. Springer, 2001, vol. 2253, pp. 11–18, 10.1007/3-540-45548-5_2. [Online]. Available: http://dx.doi.org/10.1007/3-540-45548-5_2

[11] T. K. Judge, C. Neustaedter, and A. F. Kurtz, "The family window: the design and evaluation of a domestic media space," in *Proceedings of CHI '10*. New York, NY, USA: ACM, 2010, pp. 2361–2370. [Online]. Available: http://doi.acm.org/10.1145/1753326.1753682

[12] T. K. Judge, C. Neustaedter, S. Harrison, and A. Blose, "Family portals: connecting families through a multifamily media space," in *Proceedings of CHI '11*. New York, NY, USA: ACM, 2011, pp. 1205–1214. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979122

[13] S. A. Ballegaard, T. R. Hansen, and M. Kyng, "Healthcare in everyday life: Designing healthcare services for daily life," in *Proceedings of CHI '08*. New York, NY, USA: ACM, 2008, pp. 1807–1816. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357336

[14] K. Kasugai, C. Röcker, B. Bongers, D. Plewe, and C. Dimmer, "Aesthetic intelligence: Designing smart and beautiful architectural spaces," in *Ambient Intelligence*, D. V. Keyson, M. L. Maher, N. Streitz, A. Cheok, J. C. Augusto, R. Wichert, G. Englebienne, H. Aghajan, and B. J. A. Kröse, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, vol. 7040, pp. 360–361. [Online]. Available: http://www.springerlink.com/content/263w23vr6603603t

[15] K. Kasugai, M. Ziefle, C. Röcker, and P. Russell, "Creating Spatio-Temporal contiguities between real and virtual rooms in an assistive living environment," in *Proceedings of CREATE 10*. Loughborough, UK: Electronic Workshops in Computing (eWic), 2010, pp. 62–67. [Online]. Available: http://www.bcs.org/ewic/create10

[16] M. Milman, *Trompe-l'oeil, painted architecture*. Skirá, 1986. [Online]. Available: http://books.google.com/books?id=GKySQwAACAAJ

[17] C. Overbeeke and M. Stratmann, "Space through movement. a method for three dimensional image presentation," Ph.D. dissertation, TU Delft, Delft, 1988.

[18] N. Dodgson, J. Moore, and S. Lang, "Multi-View autostereoscopic 3D display," *IEEE COMPUTER*, vol. 38, pp. 31–36, 1999. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.7623

[19] K. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita, "Multiparty videoconferencing at virtual social distance: MAJIC design," in *Proceedings of CSCW '94*. New York, NY, USA: ACM, 1994, pp. 385–393. [Online]. Available: http://doi.acm.org/10.1145/192844.193054

[20] K. O'Hara, J. Kjeldskov, and J. Paay, "Blended interaction spaces for distributed team collaboration," *ACM Trans. Comput.-Hum. Interact.*, vol. 18, no. 1, pp. 3:1–3:28, 2011. [Online]. Available: http://doi.acm.org/10.1145/1959022.1959025

[21] C. Kuster, T. Popa, C. Zach, C. Gotsman, and M. Gross, "FreeCam: a hybrid camera system for interactive free-viewpoint video," in *Proceedings of VMV '11*, Berlin, Germany, 2011.