blue-c: Using 3D Video for Immersive Telepresence Applications

Silke Lang†, Stephan Würmlin*,
Jan Borchers‡, Ludger Hovestadt†, Markus Gross*

† Chair of CAAD
Faculty of Architecture
ETH Zürich, Switzerland
e-mail: {lang,hovestadt}@arch.ethz.ch
http://caad.arch.ethz.ch/

* Computer Graphics Laboratory
Computer Science Department
ETH Zürich, Switzerland
e-mail: {wuermlin,grossm}@inf.ethz.ch
http://graphics.ethz.ch/

‡ Media Computing Group
Computer Science Department
RWTH Aachen, Germany
e-mail: borchers@cs.rwth-aachen.de
http://media.informatik.rwth-aachen.de/
blue-c: Using 3D Video for Immersive Telepresence Applications

Silke Lang¹  Stephan Würmlin²  Jan Borchers³  Ludger Hovestadt¹  Markus Gross²

¹Chair of CAAD  ETH Zurich ²Computer Graphics Laboratory ³Media Computing Group
ETZ Zurich  ETH Zurich RWTH Aachen
{lang, hovestadt}@arch.ethz.ch  {wuermлин, grossм}@inf.ethz.ch  borchers@cs.rwth-aachen.de

ABSTRACT
In this paper we describe real-time 3D video technology for immersive telepresence applications. The technology allows users to meet and collaborate in networked virtual environments in full 3D. The user can freely navigate and see himself or others from arbitrary viewing points. The photorealistic appearance of a 3D user representation allows for natural collaboration in immersive spaces and enables novel interaction techniques for telepresence. Besides interacting with the environment a user can also interact with 3D representations of himself and of remote users. To this end, we identify a design space to show how blue-c differs from current telecollaboration systems. The main contribution is 3D video as an interaction medium for immersive projection and 3D video acquisition environments. We conclude with some experiences derived from inserting and interacting with 3D representations of the users in specific virtual reality applications.

Author Keywords
Input and Interaction Technologies, Telepresence, Spatially Immersive Displays, 3D Video

ACM Classification Keywords
H.4.3 [Information Systems Applications]: Communications Applications–Computer conferencing, teleconferencing, and videoconferencing; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems–Artificial, and virtual realities; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces–Synchronous interaction.

INTRODUCTION
Finding the right interaction and communication techniques that let people communicate over distance in a natural manner has been one of the holy grails of Human-Computer Interaction. With each new technology, from written text, to graphics, to speech and video, systems have emerged that have broadened the multi modal channel between remote collaborators in order to increase the quality of their interaction with each other and their common environment.

So far, however, these systems forced users to decide between one of two options, each with their specific advantages and disadvantages:

• Videoconferencing systems today arguably provide the most natural communication channel between remote sites. With real audio and video links creating a fairly realistic overall user experience of each other and the environment. They do, however, limit the local party to a few camera feeds of the remote group (usually only one). Their angles of view may be remotely adjustable, but certainly not freely on a still frame of the remote site, or turned back even to look at the local party from an arbitrary viewpoint. Consequently, problems with eye contact are inherent to 2D videoconferencing systems, especially with multiple users on both ends.

• Virtual 3D worlds in which users are represented as avatars do not have any of these limitations. The viewpoint from which the local user or users perceive the scene can be chosen arbitrarily – they can look at other avatars (representing remote users) from any angle, or they may examine their own avatar from outside. But of course, these systems are far from providing a realistic collaborative experience. Each action of the avatar must be controlled explicitly by the user through unnatural commands or complex sensor equipment that needs to be donned first.

A key component of any system that is supposed to provide a helpful common space or an environment with ambient intelligence for multiple remote people to interact, would be a technology that breaks the either-or limitation of the solutions described above. blue-c is such a technology.

blue-c is a highly complex, custom-designed and -built virtual reality environment in which people's actual visual appearance is scanned continuously and transmitted as 3D point cloud to the remote location where it is re-rendered into an otherwise virtual 3D space for the other users to observe. Each user finds himself in an immersive virtual environment that displays a 3D representation of the common virtual space and the other user's 3D video representation.

This solution overcomes the limitations of previous systems: The visual user representation is as highly realistic as in 2D videoconferencing. Due to the point-cloud representation, it is also highly flexible and open to digital post-processing. The decoupling of the point of view (virtual camera position in the scene) and the user's representation position in the scene is a key feature. Changing the point of view of the local user, for example, is trivial. It is no problem to look, say, at the partner's back without our own representation moving at all, by moving our virtual camera position (i.e. “flying” around them), or by moving (“beam”) a user’s representation to a different spot in the scene.

As with all new media technologies, appropriate and intuitive interaction metaphors are difficult to establish for a system such as blue-c. Application scenarios are helping us to reveal the best interaction techniques. However, we have also discovered surprising new opportunities for interaction techniques hitherto unknown that blue-c enables.
We apply the design space approach used in HCI since the 1980’s, e.g. by Card et al. in their seminal paper on input devices [2], to show how blue-c differs from other systems. We close with a brief description of our future plans for research around the interactive possibilities of blue-c.

BACKGROUND AND RELATED WORK

In this section we will discuss prior work in the area of tele-collaboration systems and discuss their interaction possibilities.

There has been extensive research on Spatially Immersive Displays (SID) and Collaborative Virtual Environments (CVE) over the past decades. We refer to [9] for a detailed discussion of systems in these areas. However, all these systems do not provide simultaneous, real-time 3D video acquisition and projection. For example, TELEPORT [7] is a system with user integration by means of 2D video avatars. In the Office of the Future [12] the surrounding static background environment is scanned and depth information of the users is obtained by a structured light technique. Both systems are restricted to uni-directional operations. Kauff and Schreer [10] propose an immersive 3D video conferencing system with real-time reconstruction. They provide limited virtual view display of the user with head-motion parallax. Since they rely on a 2.5D representation by means of depth-images, arbitrary viewpoints are not possible. The National Tele-Immersion Initiative [13] uses 3D avatars as remote user representation, where arbitrary viewing points are also not possible. MASSIVE [8] is a virtual reality teleconferencing system with 3D non-realistic avatars. The main criticism on avatars is that they are acting simply as placeholders and do not enhance the quality of the communication [17]. blue-c [9] is the first system combining spatially immersive displays and 3D video for telepresence applications. This not only allows the user to freely navigate around other users and himself, but also provides the realistic behavior of the user within the virtual space.

Extensive research has been done on improving the visual and behavioral realism on perceived degree of communications. Garau et al. [5, 6] focus on avatar realism in shared, immersive virtual environments (IVE). In their studies they conclude that avatars must have certain behavioral characteristics in order to be useful. They use eye gaze animation as a specific instance of behavioral realism of avatars. Vertegaals et al. [15] investigate how visual information increases communicative bandwidth of mediated interaction. GAZE-2 [16] is a novel video conferencing system. Video images of the users are placed in a virtual environment. These images are automatically oriented towards the dialog partner each user looks at. To this end, eye-controlled camera direction is used to ensure parallax-free transmission of eye contact. Chen [3] determines in his experiments how accurately users perceive eye contact in video conferencing systems. Using video images for remote collaboration does not allow for preserving spatial relationships among participants which is paramount for certain collaborative situations, such as remote acting rehearsals [14]. However, in immersive telepresence systems where users freely navigate in space, it is difficult to provide a satisfying solution for a realistic representation of the user.

SYSTEM DESCRIPTION

In this section we give a high-level overview of the blue-c system, discuss how the user experiences the system and give a short overview of the 3D video technology. The conceptual components of the blue-c system architecture are depicted in Figure 2. For a more detailed technical discussion we refer the reader to [9].

System Overview

blue-c is a novel immersive projection and 3D video acquisition environment for collaborative work and virtual design. A blue-c portal simultaneously combines a spatially immersive display with a multi-camera acquisition stage. From multiple video streams we create a 3D video represen-
tation of users or objects in real-time. The resulting 3D video inlays are transmitted to a remote site for telecollaboration. The blue-c technology connects multiple such portals into a networked virtual environment. Its scalable design even allows connecting to portals with less sophisticated hardware such as traditional desk-side 2D video conferencing devices. In each portal spatial audio rendering enables exact location of sound sources of virtual objects. This technology also allows real-time speech transmission of participants. Figure 1 shows the two blue-c portals we built for our first experiments. One is a three-sided CAVE™-like [4] portal which is located at ETH Computing Center downtown Zurich. The PowerWall-like portal is located outside of Zurich on the second campus of ETH Zurich, approximately 10 km apart.

User Perception

The user herself enters a blue-c portal without going through any initialization phase. Depending on the configuration of the software system the user experiences either herself and/or remote collaborators, one or many. Since 3D video inlays are available as three-dimensional representations they can be easily combined with the virtual world projected onto the display. Thus, a seamless and natural integration of inlay and virtual world is perceived. Figure 3 gives an impression of the visual appearance of the perceived display.

Currently, we only extract and transmit objects in the portals, the portal itself is not represented. Since we are not using a model-based approach for extracting geometric information we can also introduce arbitrary objects into the system besides the user, e.g. a chair or a golf club. Basically everything that gets into a portal is three-dimensionally represented. To allow this separation of foreground objects from background we have to go through a learning phase. This usually has to be done once per session and takes about 10 seconds.

Acquisition and Projection

Since the blue-c portals are spatially immersive displays the user is able to work and interact with virtual worlds while being physically surrounded with a panorama of imagery. For a high degree of immersion a stereoscopic projection is inevitable. The PowerWall-portal supports stereo and mono projection, whereas the CAVE-portal only supports active stereo projection. To that end, the users inside the portal have to wear stereo glasses. This has impact on eye contact which is an important factor of human perception. Eye contact is only partly perceivable in the case of stereo projection.

The basic idea of blue-c is to hide the video acquisition components from the user so that he is not aware of it. Figure 4 shows the arrangement of cameras and projection walls for the two portals. In the case of the CAVE-portal almost all cameras are hidden behind the projection walls. The cameras can “look through” the projection walls by employing screens with switchable opacity. During a short time slot we “open” the walls for image acquisition. But since the shutter glasses are totally opaque in this phase the user is not aware of the cameras or the transparent walls. Furthermore, the inner five cameras are attached on top of the projection walls and the ceiling of the room. In the case of the PowerWall-portal the cameras are well hidden in the walls and the ceiling of the installation.

3D Video Technology

3D video objects are represented as a 3D point cloud [19]. Each point sample carries attributes such as position, color and surface normal. No connectivity between the point samples is necessary which allows for efficient and low latency transmission of 3D video objects. The surface normals allow for re-shading the 3D object according to the lighting conditions of the virtual world.
We extract three-dimensional information of the user employing a shape-from-silhouettes technique which calculates the visual hull. Since this method relies on the silhouettes alone, acquisition of multiple objects is limited to clearly distinguishable silhouettes of each object, i.e. silhouettes should not overlap in any recorded view. This means, in practice, that we can build 3D video objects of everything which is inside the portal except for objects which are too close to each other. The limit depends on the situation, e.g. two users in the same portal can be very well represented if they are approximately 30 cm apart from each other.

In addition, the 3D video technology allows to record users and objects [18] and replay these sequences from arbitrary viewpoints.

**Performance and Delay**
The performance of our 3D video engines is sufficient for processing objects with less than 30k point samples. In a typical 3D video sequence, processed at 10 frames per second we maintain between 15k and 25k points in the 3D video objects per site. Our real-time 3D video system has a system inherent latency of 3 frames. In the worst case, the round trip latency sums up to 5 frames. In experiments between the two portals at ETH Zurich the minimum latency is 200 ms and the average latency is about 300 ms. In experiments with an inter-continental connection between ETH Zurich and UC Davis we experienced latencies up to 400 ms. We refer to [19] for a detailed discussion of performance and delay.

**INTERACTION TECHNIQUES**
In this section we discuss the interaction techniques available in the blue-c system especially highlighting possibilities with the 3D video technology.

Navigation inside a virtual world and picking of objects are common tasks in most VR applications. blue-c supports common 3D motion tracking devices such as Ascension and Polhemus magnetic tracking systems. In addition, the system supports compound devices such as the Fakespace Wand™. Simple fly-motion navigation with collision detection and pre-defined or dynamic setting of viewpoints as well as basic object picking are implemented in the blue-c system.

**Arbitrary Viewpoints.** The 3D video representation introduces novel interaction paradigms not available to 2D video conferencing systems or avatar representations. Unlike 2D conferencing systems we can look at participants (represented as 3D video objects) from arbitrary viewing directions. Furthermore, as opposed to avatar systems, we can show real objects, e.g. an additional sweater, to the remote user.

**Deformations and Modifications.** It is also possible to directly interact with the 3D representation. We allow deformations and modifications of the 3D video representations. Currently, we only implemented procedural deformations, such as explosions, or sine waves through the objects. Another deformation process is “beaming” where we change the location of the object/user to a different location by floating the point samples randomly. Figure 5 illustrates this process.

**Placement and Instances.** We can freely change the location and orientation of 3D video objects in the virtual world. We can even allow arbitrary instances of 3D video objects or combine them together thus allowing overlapping of objects.

**Collision Detection and Occlusion Handling.** Conservative collision detection is easily achievable by having bounding boxes of the 3D video objects available for testing against objects in the virtual world or against other 3D video objects. Once enabled, we can move 3D video objects only in places consistent with the virtual scene. Occlusion handling is straight-forward since 3D video objects are regular three-dimensional geometry in the virtual world and thus Z-buffering does the work for us.

**Effects.** With our 3D video technology we can freeze the 3D videos, thus allowing freeze-and-rotate effects comparable to effects in the “Matrix”-movies. In addition, with the recording capabilities fast-forward and -reverse, and slow motions are also possible.

**Recognition.** Vision-based object recognition in blue-c is currently under development. Detected hand and head positions can be used as standard tracking device sensor thus allowing novel interaction, such as in the “flashlight” application: A user can point a virtual flashlight into the virtual scene. The software calculates a soft flashlight cone and only objects in this cone get illuminated. The rest of the scene is dark. Figure 6 illustrates the flashlight application.
In the future, we will look into more advanced gesture recognition with applications such as picking and navigation. In addition, we want to implement a “touchscreen” application in blue-c where the user can press virtual buttons by physically touching the projection walls. We detect these collisions by analyzing the 3D video data.

**DESIGN SPACE**

By classifying the blue-c system into several categories (see Figure 7), our goal is to define a design space for telepresence systems and applications. We propose a four-dimensional space for telecollaboration systems such as blue-c: modality, eye contact, tele-immersion, and user realism. In the following we describe each of these dimensions in more detail followed by a discussion of the categories and their place in the design space.

**Modality**

For user representation in telepresence applications, this attribute describes the appearance of the user. The attribute can be classified by the signal dimension of the user representation plus whether it is in motion or not. We distinguish between three dimensions:

1. **2D modality**: the appearance has 1D characteristic and motion, e.g., text, audio; or 2D characteristic without motion, e.g., still image, static 2D avatars (icons, comics).

2. **3D modality**: the appearance has 2D characteristic and motion, e.g., video, animated 2D avatars; 3D characteristic without motion, e.g., static 3D avatars; or 3D characteristics from predefined locations, e.g., 3D representations with limited viewpoint possibilities, i.e., 2.5D representation of the user.

3. **4D modality**: the appearance has motion and 3D characteristic from arbitrary locations; e.g. 3D video, animated 3D avatars with arbitrary viewpoints.

**Eye Contact**

This attribute indicates whether a system provides face-to-face communication between one or more participants. Eye contact is experienced when the people look into each other’s eyes. In visual communication people also use their eyes to express themselves.

**Tele-Immersion**

Tele-immersion refers to how a system can make users feel like they are experiencing an alternate reality and not just merely observing it. Moreover, this technique enables users at different physical locations to meet and collaborate in a simulated, shared environment in real-time. The degree of immersion depends on one hand on the feeling of presence, of “being there”, surrounded by the space and capable of interacting with all available objects. On the other hand it depends on the perceptual reaction whether the communication between the users is experienced like a real face-to-face communication.

**User Realism**

In telepresence systems the users are represented with different kind of media. User realism describes whether a user has a synthetic or a real representation in the virtual world. There are two major approaches: synthetic and real.

---

**Figure 7:** Design space for tele-collaboration systems

![Design Space Diagram](image-url)
One approach is the identification of handles, aliases or nicknames in text-based virtual communities, and avatars, also called characters, players, virtual actors, icons, or virtual humans in virtual environments. We call these synthetic representations.

Another approach is photorealistic representation of the users such as 2D video images or 3D reconstruction of real persons and objects employing vision algorithms. We call these real representations.

Telecollaboration categories

In the remainder of this section, we will briefly describe a subset of interaction and communication techniques that let people communicate over distance.

In text-based environments (1), such as chat rooms or multi-user domains (MUDs), the user representations have a 2D modality. Audio-based environments (2), e.g. telephones, have a real user representation. The appearance of the user has a 1D characteristic and motion. Both systems do not allow for eye contact between the users. 2D avatar or comic representations (3) as used in The Place (http://www.theplace.com) have a 3D modality, e.g. the avatars can change their expressions. MASSIVE [8] and ACTIVE-WORLDS [1] for example are immersive teleconferencing systems with 3D avatar user representations (4). These systems have synthetic user representations with a 4D modality. Eye contact between the users is not possible. Videoconferencing systems (5), for instance GAZE-2 [16] and desktop video conferencing (DVC) provide the most natural communication channels between remote sites using 2D video. Eye contact with multiple users is still an inherent problem. Immersive tele-video systems (6), such as the Office of the Future [12], the National Tele-Immersion Initiative [13] and 3DVC/VIRTUE [10] provide real representations of the users with a 3D modality. A significant difference to immersive 3D video systems (7) is that they only allow for limited viewpoint trajectories. The blue-c system provides an immersive 3D video representation of the users. The actual visual appearance of the users is rendered in the immersive space of the remote users. Therefore, the users have a real, 4D modality representation. As opposed to other systems blue-c provides arbitrary viewpoints and eye contact between multiple users.

APPLICATIONS AND DISCUSSION

To demonstrate our 3D video technology as novel interaction medium for telepresence we implemented a set of prototype applications. The applications present the different interaction techniques (described in the section Interaction Techniques), the interaction with oneself and the communication between physical distant users. This allows us to test our interaction techniques and evaluate them with different kind of users. We describe two applications MIRROR:3D and IN:SHOP [11] in more detail in the remainder of this section.

MIRROR:3D

In the application MIRROR:3D the user is able to experience himself in a virtual world in real time. The virtual space could be for example a virtual fashion show, where the user experiences herself on a virtual catwalk or dance studio where the user experiences his movements from all viewing directions. The MIRROR:3D is similar to a regular 2D mirror. The significant difference is the possibility to freely navigate and see oneself in full 3D from arbitrary viewing directions. For example, it is possible for a dancer to see his back view. A golf player could move his representation into his remote teacher. This allows him to check his swing immediately. Figure 8 gives an impression of this application.

Figure 8: MIRROR:3D: The user experiences herself and interacts with the system

Test Plan

The users were asked to enter the CAVE-like portal and experience the possibilities of the 3D video technology. Before starting the test run, the test users are introduced in the application and explained how the 3D mirror works. The users were explained how to navigate within the virtual world and how they can fly around themselves. The overall mount-time is very small. Speech and gestures have no learning-time because the user does not need to engage any special equipment. He can act in his natural and intuitive way. The 3D mouse has a small learning-time since it functions like the traditional computer mice do. The stereo glass also requires a very small mount-time since it only must be put on. We ask them to explore the possibilities and to feel free in interacting with the system.

User Experience

The primary variable of interest is to evaluate the user’s interaction with the system and the possibilities he explores to do so, divided into the following three indicators, similar to [5]:

1. **Learning-time**: The extent to which effort is necessary to use the system.
2. **Latency**: The extent to which latency is acceptable / tolerable for the user interaction.
3. **Fun**: The extent to which users enjoyed experiencing the application.
Preliminary Results

We had 15 persons with different backgrounds visiting our blue-c e system and taking part in our experiment. All of them were technically skilled, nearly two thirds stated to use telepresence systems. Half of them already had experiences with virtual reality systems and immersive environments.

The reponse of the variables described above were obtained by means of a questionnaire with each participant. We noticed that all users were hesitant and reserved using the application in the first minute. After some shy movements they started interacting with themselves. All users started with waving and to moving around. Surprisingly, while flying around them and watching them from behind all users started looking over their shoulders trying to see their real backs. After a while most of the users got used to navigating in the virtual environment using the 3D mouse. The users who already were familiar with navigating in immersive VR applications zoomed their representation in an out. 20% of the users explored the possibility of taking of and putting on some clothes, e.g. sweater or jacket. They were fascinated by being able to see the changes on their representations. All users tried out the latency of the system. They either moved up one arm very fast or jumped up and down. They felt that there is some latency, but found that it is tolerable and does not minimize the experience.

The users were very enthusiastic about interacting with themselves in full 3D. They appreciated seeing their actual visual appearance in 3D in the virtual space. The users who were not familiar with navigating in immersive virtual worlds found that they have to get used to the 3D mouse. All participants were impressed by the fact that they were able to see their backs and flying around themselves. However, all of them were fascinated how easy the system is to be used and how small the mount-time is.

The overall response was that all users liked this application and enjoyed seeing themselves from arbitrary directions in real-time.

IN:SHOP

Another application shows how 3D video technology, communications, and architecture can be combined in designing a novel, enhanced immersive shopping environment. We call this application IN:SHOP. It uses our two portals at ETH Zurich to implement a solution for distributed shopping. Figure 9 shows the application in action.

Virtual Space

IN:SHOP allows a customer to meet a sales assistant in a virtual shop. Together they browse through the content depending on the customers’ requirements. If the customer has indicated interest in certain items, the sales assistant presents them to the customer using a variety of traditional and VR based media. Traditional media vary from simple text to images. VR based media are ranging from CAD data to 3D video. The shopping process for the customer would see the object gradually becoming more detailed and interactive towards the agreement for a purchase. Before agreeing to buy a product the customer would be able to interact with a 3D representation of the real object.

We implemented this concept for a fashion shop (Fashion Shop) and a car selling application called CarShop.

Test Plan

Before starting the real test run, the test persons are introduced to the functionality of the system and the shopping environment. The use of natural speech and gestures as well as the common navigation tasks using a standard six degree of freedom mouse with a small joystick and 3 buttons (Fakespace WandTM) is explained. In order to assess whether the presence of a real-time 3D representation improves the quality of communication, we need a task in which participants would benefit from having visual feedback. Therefore, we use our CarShop application. In CarShop the customer and sales assistant configure a car. Potential buyers interactively configure and experience 3D representations of the car they are interested in. The sales assistant advises the customer comparable to the traditional situation where all participants are present.

It is pointed out that this is not an acting task and that all participants should be themselves and should feel free to improvise. This helps us to find out if in same-room interaction, factors such as interpersonal distance, posture, and gesture might have an influence.

User Experience

The fundamental variable of interest is the suitability of 3D video for user interaction. We expect that 3D video improves the communication within telepresence applications. The questionnaires are subdivided into the following five indicators, similar to [5].

1. **Eye contact**: The extent to which the shopping process was experienced as being like a real face-to-face counseling.
2. **Immersion**: The extent to which the test users experienced the virtual shop and felt as a part of it.
3. **Presence**: The extent of presence of the remote sales assistant, i.e. meaning the sense of presence of the remote user.
4. **Communication**: The extent to which the interaction and communication were comparable to the traditional situation, and the extent to which the experience was enjoyable.
5. Learning-time: The extent to which effort is necessary to use the system.

Preliminary Results

The same users testing our MIRROR:3D application were asked to evaluate the interaction possibilities on the basis of our CarShop application.

We noticed that all of the users tried more than once to grab 3D objects within the scene or to touch the 3D representation of the remote user. All users found it very important for the counseling process that they not only hear the sales assistant but also see his gestures. They felt that this greatly enhances the sense of presence and reality within the virtual shop. They liked being able to point on information displayed in the virtual shop. All users found that they were able to talk to the remote user like in the traditional situation where both are in the same physical space. They all enjoyed configuring their personal car with a remote sales assistant. Even though our stereo glasses minimize real eye contact between the customer and the sales assistant, nearly all user experienced the shopping process closely to a face-to-face client counseling. They considered to use smaller glasses comparable to the size of sun glasses. Users familiar with other telepresence systems liked this concept because they felt that the 3D representation gave them a better understanding of the remote user.

Although the users were generally enthusiastic about the potentials of the system, they felt that some improvements can be done. Half of the users would appreciate using the screens as virtual touch screens. Some of the users proposed to use PDAs as input devices. A third of the users wish to be able to put the remote user their real jacket on.

Overall, the users we interviewed were enthused and felt that the concept would be helpful in increasing the realism in immersive telepresence applications.

CONCLUSION AND FUTURE WORK

In summary, 3D video has proved to be a flexible prototype interaction medium for telepresence applications. This paper described a number of interaction techniques to enhance communication in distributed virtual worlds. An important aspect of our 3D video interaction techniques is the use of real-world interaction metaphors. We hope that real-time 3D video will help us and others to further explore and systematic study interaction techniques for immersive projection and 3D video acquisition environments.

In the future, we want to design new applications with the 3D video technology especially in the field of collaborative virtual design and architecture. Furthermore, we will perform a profound usability study to classify application scenarios and to quantitatively measure the benefit of 3D video for immersive telepresence applications. In addition, manipulation of point clouds and their metaphors and vision-based gesture recognition are interesting areas of future work.

REFERENCES