Exploring Augmented Live Video Streams for Remote Participation

Michael Wittkämper
Fraunhofer FIT
Schloss Birlinghoven
53754 Sankt Augustin
Germany
michael.wittkaemper@
fit.fraunhofer.de

Jan Ohlenburg
Fraunhofer FIT
Schloss Birlinghoven
53754 Sankt Augustin
Germany
jan.ohlenburg@
fit.fraunhofer.de

Irma Lindt
Fraunhofer FIT
Schloss Birlinghoven
53754 Sankt Augustin
Germany
irma.lindt@fit.fraunhofer.de

Jan Herling
Fraunhofer FIT
Schloss Birlinghoven
53754 Sankt Augustin
Germany
jan.herling@fit.fraunhofer.de

Wolfgang Broll
Fraunhofer FIT
Schloss Birlinghoven
53754 Sankt Augustin
Germany
wolfgang.broll@
fit.fraunhofer.de

Sabiha Ghellal
Sony NetServices GmbH
Kemperplatz 1
10785 Berlin
Germany
sabiha.ghellal@eu.sony.com

CHI 2007, April 28 – May 3, 2007, San Jose, California, USA.
ACM 978-1-59593-642-4/07/0004.

Abstract
Augmented video streams display information within the context of the physical environment. In contrast to Augmented Reality, they do not require special equipment, they can support many users and are location-independent. In this paper we are exploring the potentials of augmented video streams for remote participation. We present our design considerations for remote participation user interfaces, briefly describe their development and explain the design of three different application scenarios: watching a pervasive game, observing the quality of a production process and exploring interactive science exhibits. The paper also discusses how to develop high quality augmented video streams along with which information and control options are required in order to obtain a viable remote participation interface.

Keywords
Augmented reality, augmented video streams, video content/communications

ACM Classification Keywords
H5.1. Information interfaces and presentation (e.g., HCI): Multimedia information systems: Artificial, augmented and virtual realities; Video

Introduction

Augmented Reality (AR) systems as defined by Azuma [1] overlay the real with the virtual, are interactive in real-time, and registered in 3D. A major benefit of AR systems (as compared to Virtual Reality systems) is that they support presenting information in the spatial context of the physical environment. This allows for a better understanding of the respective data making AR systems suitable for several application areas including education and training, manufacturing, as well as games.

One drawback of AR systems is that they are only available for a limited number of users since each user requires specialist and expensive hardware such as an accurate head tracking system, a special display and high-performance rendering hardware. In contrast augmented live video streams overlay virtual artefacts onto the real environment which can be shared by many people simultaneously. Moreover, they support remote participation via different, non-specialized computing hardware (typically only a web browser or a video streaming client is required).

In this paper we present the design and development of augmented live video streams. We start by introducing some design considerations for augmented live video streams and by explaining the development process. A major part of this paper is the description of three application scenarios, which illustrate the range of uses for augmented video streaming user interfaces and how they are accepted and employed by users. We conclude the paper with further observations and a few recommendations.

Design considerations

In order to design augmented video streaming applications for use in remote participation, we identified several issues that need to be considered:

- mobility and number of employed cameras
- support for video stream control
- type of supporting information
- availability of audio channel
- supported degree of participation
- number of supported clients
- available bandwidth
- desired video properties (frame rate, resolution, latency)
- suitable codecs and video formats

The mobility aspect encompasses static, pan/tilt/zoom, robot-mounted, and head-mounted cameras. Video stream control is the process of switching between several streams or controlling the pan, tilt or zoom. The number of employed cameras, their mobility, as well as the supported degree of video stream control influences the type of camera which must be chosen in order for the augmentation to cover the required area of the real environment. Apart from the video stream, further information, e.g. in the form of text, graphics or optional audio channels may be necessary in order to support the desired degree of participation via the remote user interface. Finally, the number of supported clients, the available bandwidth, and the desired video properties have an impact on the system configuration and especially on the choice of codecs and video formats.

Realization

In our system (which is part of the AR/VR Morgan Framework[2]), augmented live video streams are created in the AR system's visualization component and distributed to various clients through an IP network (see Figure 1).

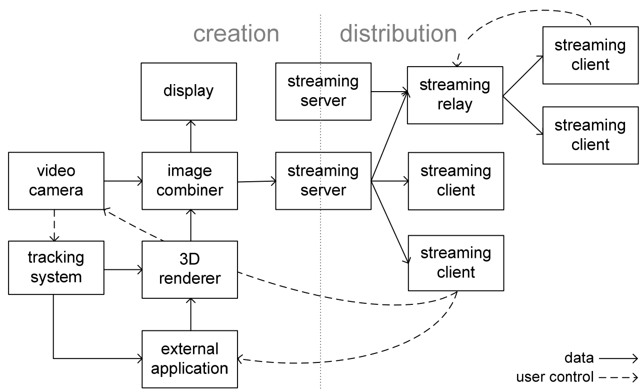

Figure 1. Augmented video streaming components for stream creation and distribution including data and user control flow.

The AR scene, optionally controlled by a head tracking system and an external application, is rendered and combined with a video background. To eliminate camera distortion and to adjust the camera's field of view and pose to that of an AR display, this background is created by mapping the camera image to the inner surface of a sphere. Care has to be taken that the camera's field of view is both large enough to cover the whole background and small enough so that the remaining pixel resolution on the background is sufficient. In case the camera is part of a video see-through display (e.g. an HMD) the frame can be shown in the display. For optical see-through displays a second render pass without the video background is required. The combined frame is then compressed and transmitted over the network.

RTSP/RTP [3] was chosen as the stream transport protocol since it is supported by a wide range of media players and streaming servers. The challenge in augmented live video streaming is to encode and transmit each frame without limiting the real-time behavior of the system. Since the bottleneck lies in the frame compression stage, the choice of the right encoder is of vital importance. MPEG4 encoders were chosen as they ensure an acceptable CPU rate, frame rate and quality as well as data size, but other encoders are also supported. By implementing the streaming server as a DirectShow filter we gained a high degree of flexibility. The complete process of video streaming can be carried out within a DirectShow graph, making it easy to test and exchange different MPEG4 encoders. Additionally it forwards the encoded frames to other filters, e.g. for recording the stream to a file.

In case several clients are connected simultaneously or they are located in different networks with differing bandwidths (e.g. intra- and internet) streaming relays are employed to serve the individual clients, to reduce size and frame rate, and to support different protocols such as HTTP. Application specific augmented live video streaming user interfaces are implemented outside the framework offering a number of features such as pan/tilt/zoom camera control or choice between several video streams.

Application scenarios

The following three application scenarios illustrate how augmented live video streams were employed for remote participation in augmented environments.

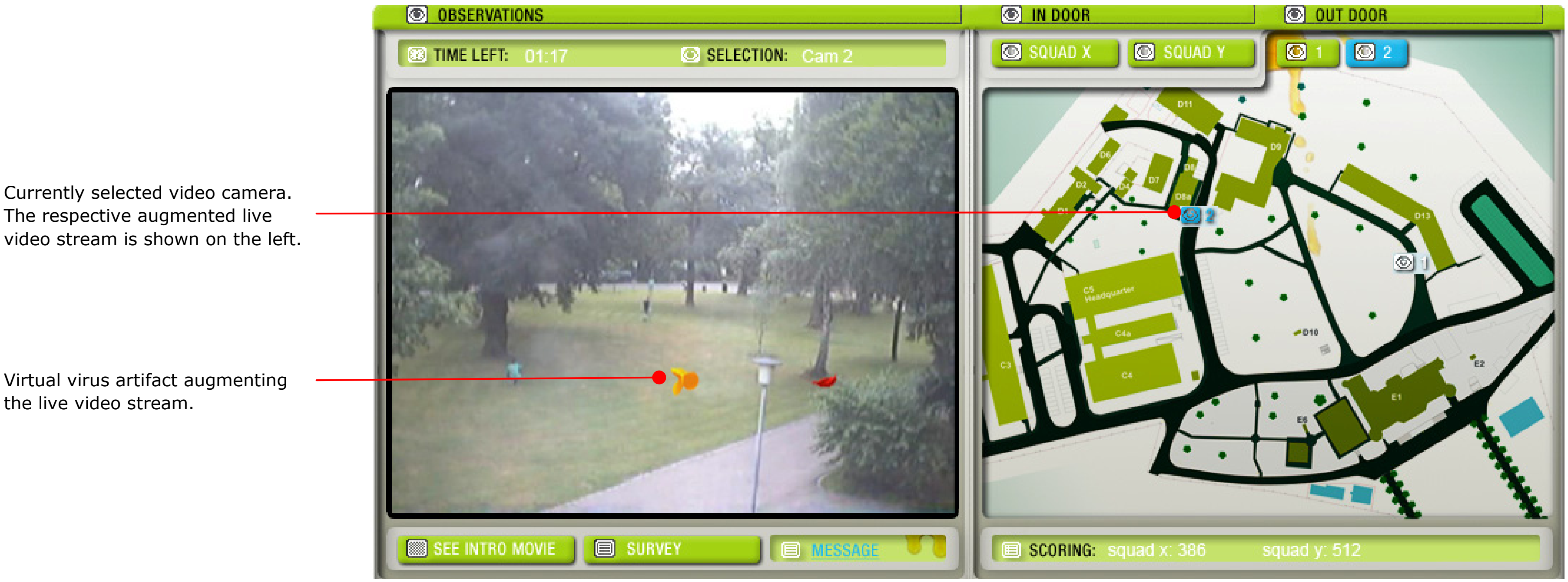

Figure 2. Epidemic Menace web-based spectator interface with augmented live video stream (left) and a site map (right).

Pervasive Game Spectator Interface

For the pervasive cross media game Epidemic Menace we implemented a web-based spectator interface based on augmented live video streams (see Figure 2). During the game play, which lasts typically for about three hours, players can be watched while they are walking around on a virus-contaminated campus, observing and trying to destroy the viruses. For the spectator interface we used four low-cost cameras and positioned two of them indoors and two outdoors. The spectator can switch between the different camera views using a map on which the cameras are located. The indoor cameras show the indoor players observing the game area and directing the outdoor players. The outdoor cameras show large parts of the game area overlaid with animated virtual viruses that move and replicate. We tested the spectator interface over four game sessions during a two day period and about 30 spectators took part. Spectators had to register via a web form and we asked them to fill out a questionnaire afterwards. Spectators said that they liked the fact that they could see an augmented view of the game area and players chasing the virtual viruses. They also noted that an audio connection and more background information are crucial to following the game play and even suggested a presenter, similar to those found in TV sports broadcasts to explain what is currently happening.

Observing the quality of a production process
In the production and quality management process of large scale construction parts a number of individual experts must be consulted. However these experts are not necessarily available at the production location. To minimize the experts' equipment, a live video stream of a local service person's augmented view is transmitted to the expert. This lets the expert see the augmented part information (which the local service person is viewing) alongside their current view (see Figure 3). Additionally the expert may direct the service person towards the critical areas of the part so as to have a closer look. The only requirements on the expert's side are a standard streaming video viewer and a VoIP tool.

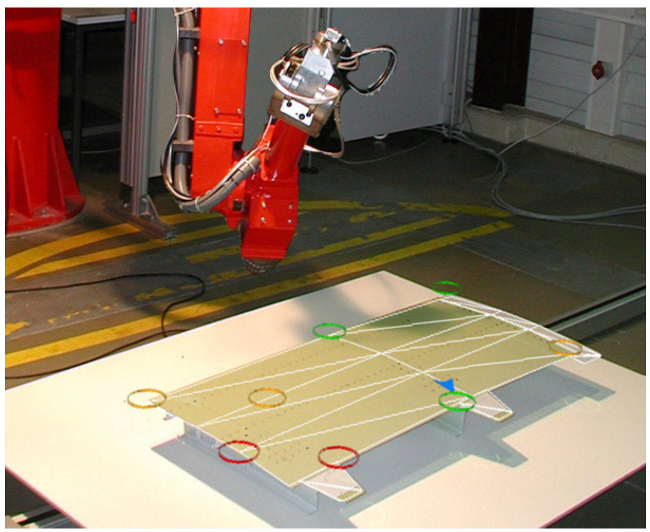

Figure 3. Service person's augmented view on questionable soldering points on airplane part.

In this application scenario a high image quality is essential whereas the frame rate is of minor importance. Directing the service person to a particular viewpoint turned out to be error-prone and sometimes cumbersome. Therefore it was suggested to mount the camera on the robot arm and provide the expert with a method of controlling the robot arm.

Exploring interactive science exhibits
In this application scenario students were equipped with mobile AR systems while exploring exhibits in four science centers across Europe. By making the invisible visible AR technology allows direct perception of complex phenomena and underlying physical models and thus makes the understanding of these concepts easier.

Two kinds of augmented streaming scenarios are supported. In the first scenario, students, teachers and supervisors located at the same exhibit as the student can monitor their view on a local projection screen or display. As both the student and the teacher are at the same location and are able to interact with one another the level of participation is very high Moreover, as only the local area network is used the available bandwidth is comparably high and the video quality is mainly dependent on the selected video camera. As only off-the-shelf webcams are utilized at this time, the effective pixel resolution is about the same level as QVGA; this results in the image quality of the image portion being rather low. No relays are required. From the user's perspective a latency of about one second proved to be acceptable.

In the second augmented streaming scenario, students in several remote classrooms spread across Europe follow the student's or a museum staff member's activities through a live video stream (see Figure 4).

They can also communicate with them through bidirectional audio connections. Here the degree of participation is reduced; remote users can either instruct the AR user to simply navigate to a specific location and look at a certain item, or suggest how to perform an experiment. However, being able to participate was well received by the students.

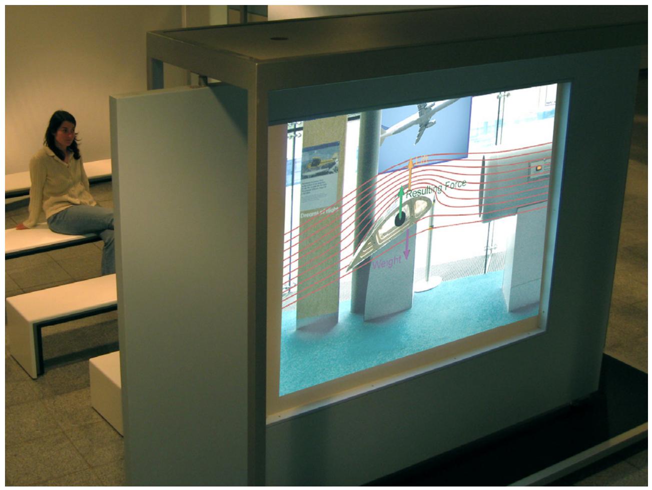

Figure 4. Student following a classmate's view of an interactive exhibit (real aerofoil augmented with virtual airflow and force vectors).

Conclusion

Our work was motivated by the desire to find out if it is possible to facilitate an AR-like experience with augmented live video streams and to which extent augmented video streams are viable for remote participation. We found that video streams are a "narrow" information channel and sometimes require a presenter to explain the content shown on the video stream. In addition, we observed that the quality of the video (including the physical environment and the virtual artifacts) plays an important role for the user acceptance. If either the physical environment or the overlaid virtual artifacts are shown in a low-resolution or with a small update rate, users will not be willing to watch the augmented video stream for a long time. However, the desired quality differs between application scenarios. In case the user moves fast the augmented video stream delivers shaky pictures and tracking delays are revealed as registration errors. With the help of a fixed camera (maybe remote controlled) a steady 3rd person view can be obtained. Regarding the AR-like experience we found that the level of camera control and the responsiveness of the system are important factors. In general, users were quite impressed by the augmented view they received via the augmented video stream.

Acknowledgements

Parts of the work described in this report were performed within the projects IPerG (004457) and CONNECT (507844), partially funded by the European Commission in FP6. We thank the aforementioned projects partners, especially Jens Maiero, Andreas Maier, Christos Theodorakakos for installing and operating augmented video streaming relays.

References

[1] Azuma, R.T. A Survey of Augmented Reality. In *Presence: Teleoperators and Virtual Environments 6*, 4 (1997), 355-385.

[2] Ohlenburg, J., Herbst, I., Lindt, I., Fröhlich, T.; Broll, W. The Morgan Framework: enabling dynamic multi-user AR and VR projects. In *Proc. VRST 2004*, ACM Press (2004), 166-169.

[3] Perkins, C. *RTP Audio and Video for the Internet.* Sams, 2002.