Interactive Content Creation with Virtual Set Environments

Jens Herder
Virtual Sets and Virtual Environments Laboratory
Department of Media
Duesseldorf University of Applied Sciences
Josef-Gockeln-Str. 9 D-40474 Duesseldorf
Germany
voice/fax: [+49](211) 4351-310
e·mail: herder@fh-duesseldorf.de
www: http://vsvr.medien.fh-duesseldorf.de

Abstract

Digital broadcasting enables interactive TV, which presents new challenges for interactive content creation. Besides the technology for streaming and viewing, tools and systems are under development that extend traditional TV studios with virtual set environments. This presentation reviews current technology and describes the requirements for such systems. Interoperability over the production, streaming, and viewer levels requires open interfaces. As the technology allow more interaction, it becomes inherent difficult to control the quality of the viewers experience.

Keywords: virtual studio, virtual sets, interactive TV, interactive content creation

1 Introduction

Virtual set environments are used for creating content for interactive TV. The mixture of real and virtual images enables affordable productions while delivering a high visual quality. Virtual set environments do not require large real sets because the presented space can be expanded virtually.

Interactive TV with virtual studios based on the use of set top boxes and MPEG 2 is presented in [Von01]. Necessary application data is encapsulated using Digital Storage Media- Command and Control (DSMCC) specification as part of MPEG 2. Feedback into the studio is transmitted using the Internet (via sockets).

Figure 1 shows the layout of a virtual set environment for live mixed media streaming. Compared to an approach using MPEG 2, which transmits a composite video signal, all data is composed at the client site. This allows having 3D interactive objects in the stream. The process of camera tracking introduces a delay. The video and audio delay line (here as a separate unit) allows synchronized processing of the data for preparing the broadcast content. In case of offline production without realtime preview such a unit can be omitted. Special workstations designed for virtual studios (e.g., ORAD DVG-10) contain video and audio delay lines as well as a chromakeyer. Additional tracking includes capturing the position of actors for resolving occlusion with virtual objects. Such tracking or additional input devices (e.g., stylus) can be used for interactive control of the content. Additional video sources are included in the scenes a video textures.

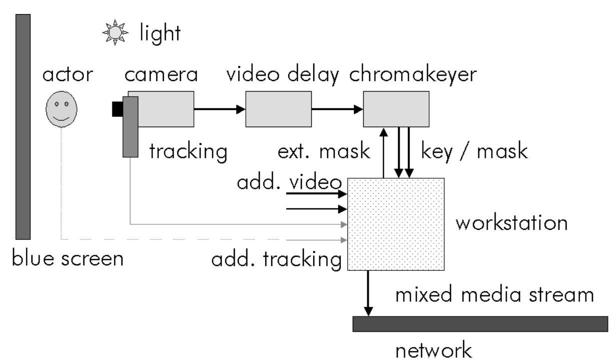

Figure 1: Virtual Set Layout for Live Mixed Media Streaming

2 Open interfaces

The interoperability between different applications of a production environment is either established by documented file formats or plugins using an application programmer interface. The plugins have the advantage of tighter integration and accessing functionality of another application using a common user interface. For the system design of virtual studios a main (host) application can be chosen. This could be the modeling and animation software with a realtime renderer plugin (e.g., ORAD Cyberset NT uses 3D Studio Max as host application [ORA01]). Other developments prefer to run the realtime renderer including live interaction control as a standalone applications with plugins for reading data from other applications like the modeling. Some virtual studio applications require that all interactions and animations are specified within the system and cannot or only partly be imported from other applications. Animations are mostly based on keyframes. Support for importing expressions and rules for enabling physical-based animation is desired, but not common available. This limits the interoperability between tools. Promising is the use of XML for data exchange or open standards like VRML [BCM97].

A realtime renderer with interaction control of a virtual studio has multiple interfaces to devices and applications as shown in Figure 2. The renderer reads geometry data and animation (based on keyframes) via a modeling & animation interface. A real camera controls a virtual camera by means of tracking. The required parameter like position, orientation, focus and zoom are read using a camera interface. Interaction devices like a stylus can be used to control virtual objects (e.g., Brainstorm Multimedia [Mul01] demonstrated a virtual umbrella controlled by a stylus at the IBC 2001). Such devices are supported by interfaces which provide device abstraction. Those 3D interfaces are not integrated on the level of the operation systems yet, while development tools do support them. Moderator and object tracking has several applications. The basic use is for resolving occlusions (e.g., when is an object in front of the operator). It can be also used for generating shadows of the moderator if not a real shadow is used. This can be implemented in software by rendering an approximated shadow. Another approach uses an additional camera at the direct light source, following the moderator. This image is than mixed to the composite as a dark shadow. An interface for a streaming module enables new applications like clickable video and webcasting. Via a database interface data be imported in realtime and used for visualization (e.g., election results, weather data.) Feedback from the audience in the studio or at remote distance can be incorporated into the system to drive interactive programs. Many TV programs use repeatedly the same virtual set with same scripting and prefabricated interaction paths. Those can be provided by external libraries and can be part of soft sets.

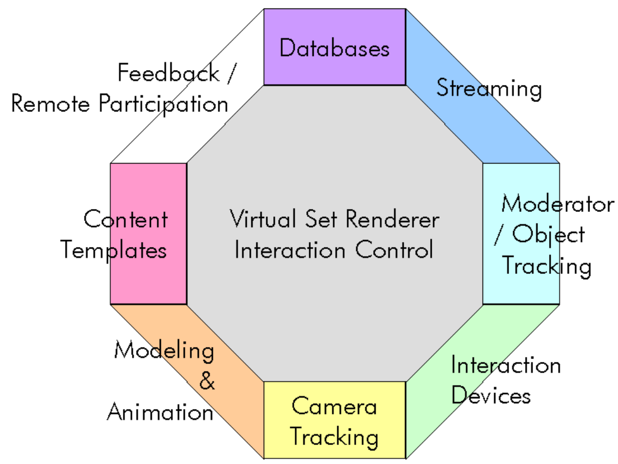

Figure 2: Multiple interfaces

3 Low-bit rate with high presentation quality for business TV

Business TV for corporate information systems has special requirements, which can be ideal supported by virtual set environments with a streaming extension. The content creation should be easy, efficient, and affordable, while special visual effects are not required. Corporations have large intranets with PCs installed, which makes it affordable to deploy TV over IP. A PC has a larger screen resolution than a regular TV. When scaling up compressed video streams of text and graphics, the visual quality is low. Besides speeches for strategic information and motivation as support for corporate identity, information like text and graphics (e.g., business charts) are distributed. The text and graphics should have a high visual quality, because they carry the main information. The technology of broadcasting a mixture of video, text and graphics based on standards like MPEG 4 [Koe01] has the advantage that the bit stream can be low but text and graphics have high quality and can be scaled up to any resolution.

4 Spatial auditory emancipation of the audience

With spatial auditory displays allowing realtime composition of the experience, the audience can emancipate itself spatially from the prefabricated (spatially composed) content. Virtual concerts are possible with the multiple audio windows [CK95][Coh94], in which the listener can rearrange the (virtual) sound sources and sinks, creating an imagery of his own choice. Virtual concerts are also presented in virtual environments [CK99] allowing also the listener to take active part in the concert by substituting one source.

Creating content for such systems requires that all voices be recorded separated and nearly anechoic. With numerous audio channels available and consumer spatial auditory display becoming available, vitual concerts can become another common listening experience [Her00]. Virtual studios have all position and orientation data for virtual and real objects (via moderator tracking or interaction devices). This can be used to drive the virtual sound imagery directly. Manipulation of the soundscape and spatial content creation of the audio elements can be supported with visual tools presented in [Her98]. All operations should be verified regarding their resource use at the client side as well network limits.

The MPEG 4 standard [Koe01] could ensure the interoparatibility of content production and content viewing. The MPEG 4 standard supports spatialization of audio streams, including sound synthesis [SVH99].

5 Conclusion

Open standards for interfaces between the building blocks of a virtual studio would allow easier integration and would enable the production site to select the best tools for their specific needs. A standard for lense calibration data, including its implementation, is one of many requirements. As long as the technology improves with such speed and the number of users (for content creation) is limited, proprietary interfaces will dominate the market. For the consumer site open standards are well established.

Virtual set environments have been described for creating interactive streams. The video from the real scenes are transmitted either as composite or screen alligned video texture. Image-based rendering methods might give even more freedom of movement for the viewer. A prototype of viewpoint-based walkthrough system for real scenes using sprites with depthmap is presented in [SNR01]. Future developments will enable more realism while expanding the degree of interaction.

References

[BCM97] Gavin Bell, Rikk Carey, and Chris Marrin, ISO/IEC *14772-1:1997: The Virtual Reality Modeling Language (*VRML97*)*, 1997, `http://www.vrml.org/Specifications/VRML97/`.

[CK95] Michael Cohen and Nobuo Koizumi, *Audio Windows for Virtual Concerts II: Sonic Cubism*, Video Proc. ICAT/VRST: **I**nt. **C**onf. **A**rtificial Reality and **T**ele-Existence/Conf. on **V**irtual **R**eality **S**oftware and **T**echnology (Makuhari, Chiba; Japan) (Susumu Tachi, ed.), ACM-SIGCHI (TBD), SICE (**S**ociety of **I**nstrument and **C**ontrol **E**ngineers), JTTAS (**J**apan **T**echnology **T**ransfer **A**ssociation), and NIKKEI (**Ni**hon **Kei**zai Shimbun, Inc.), November 1995, p. 254.

[CK99] Hartmut Chodura and Arnold Kaup, *A Virtual Environment for Interactive Music Reproduction*, Proceedings of IFIP TC/WG5.10 and CSI International Conference on Visual Computing 1999 (S.P. Mudur, D. Shikhare, J.E. Encarnacao, and J. Rossignac, eds.), International Federation for Information Processing (IFIP), February 1999, `http://www.igd.fhg.de/igd-a9/research/audio/index.html`, pp. 95–100.

[Coh94] Michael Cohen, *Conferences, concerts, and cocktail parties: Besides immersion*, JMACS: Proc. **J**apan **M**usic **A**nd **C**omputer Science **S**ociety Meeting (Musashino, Tokyo), February 1994, pp. 17–26.

[Her98] Jens Herder, *Tools and Widgets for Spatial Sound Authoring*, Computer Networks & ISDN Systems **30** (1998), no. 20-21, 1933–1940.

[Her00] ______, *Interactive Sound Spatialization - a Primer*, MM News, University of Aizu Multimedia Center **8** (2000), 8–12, (Japanese).

[Koe01] Rob Koenen, ISO/IEC *jtc1/sc29/wg11 n3156: Mpeg-4 Overview - (singapore version)*, March 2001, `http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm`.

[Mul01] Brainstorm Multimedia, *Brainstorm studio*, Product description, 2001, `http://www.brainstorm.es/`.

[ORA01] ORAD, *Cyberset* NT, Product description, 2001, `http://www.orad.co.il/virsets/index.htm`.

[SNR01] Jürgen Stauder, Yannick Nicolas, and Philippe Robert, *A Viewpoint-based Walkthrough System for Real Scenes*, Fourth Int. Conf. on Human and Computer (Aizu-Wakamatsu, Japan), University of Aizu, September 2001.

[SVH99] Eric D. Scheirer, Riitta Väänänen, and Jyri Huopaniemi, *Audiobifs: Describing Audio Scenes with the MPEG-4 Multimedia Standard*, IEEE Transactions on Multimedia **1** (1999), no. 3, 237–250.

[Von01] Wolfgang Vonolfen, *Using Virtual Studios for Interactive Digital* TV, Fourth Int. Conf. on Human and Computer (Aizu-Wakamatsu, Japan), University of Aizu, September 2001.