Creating a Live Broadcast from a Virtual Environment

Chris Greenhalgh, Steve Benford, Ian Taylor,[1] University of Nottingham, UK

John Bowers[2], Royal Institute of Technology (KTH), Stockholm, Sweden

Graham Walker[3], BT Laboratories, UK

John Wyver[4], Illuminations Television, UK

Abstract

"Inhabited Television" combines multiuser virtual environments with television, so that online audience-members can participate in TV shows staged in a virtual world. It is presented simultaneously both to conventional passive viewers and to on-line participants. In many cases it benefits from being broadcast live. This paper is based on our fourth major experiment with Inhabited TV, a live virtual game show called "Out Of This World". For this event we adopted non-automated approaches to camera control and mixing to allow an exploration of appropriate forms of presentation for inhabited television. We describe the techniques which were used to create and enhance the live video output which was produced during the show: appropriate world design; dynamic constraints on participant movements; and a performance-oriented virtual camera control interface. This camera control interface includes explicit support for a range of spatial and temporal control styles. We also give evaluative feedback on the camera control interface and the event's (television-based) approach to mixing and directing, drawing on a social scientific field study conducted on-site during the preparation for, and performances of, the show.

CR Categories and Subject Descriptors: I.3.8 [Computer Graphics]: Applications; I.3.6 [Computer Graphics]: Methodology and Techniques – Interaction Techniques; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Virtual Reality.

Additional Keywords: Multi-User, Networked Apps, Video, Viewpoint control.

1 INTRODUCTION

Multiuser virtual environments are shared 3D audio-graphical worlds in which distributed users can communicate, interact, compete and undertake other tasks and activities. We are interested in the use of these environments to create "inhabited television"; this combines multiuser virtual environments with television, so that online audience-members can participate in TV shows staged in a virtual world. We believe that this may become the "killer application" for multiuser virtual environments. Inhabited TV goes beyond the capabilities of traditional broadcast television (and more recent interactive TV systems) by allowing the online audience to interact socially with one another and to become directly involved in the content, e.g. becoming a participant in a game show or debate. Inhabited television offers broadcasters access to new formats, structures and audiences; it offers existing virtual environments a motivation for participation, an impetus for development, and a route to potential exploitation.

One of the distinctive characteristics of inhabited television is that it is presented simultaneously both to conventional passive (or interactive) viewers and to on-line participants, i.e. "inhabitants". The online participants may interact with the virtual environment through a networked PC or future set-top box. The viewers watch a traditional (or digital or interactive) broadcast television program.

There are two major benefits to making an inhabited television broadcast live rather than pre-recorded:

- interactive TV viewers can directly influence events in the virtual world as the program unfolds, for example by voting for particular changes or outcomes; and
- during the program viewers can choose to become inhabitants and vice-versa (assuming that they are using suitable equipment to view the program).

Consequently, one of the key requirements for Inhabited TV is to create live broadcasts from within a virtual environment. This paper describes the approaches and techniques which we have used to create an effective live broadcast (into a theatre space) from a multiuser virtual environment during a unique public experimental Inhabited TV show, "Out Of This World" (OOTW). This was staged at isea98revolution, part of the International Symposium of Electronic Arts, at the Green Room, Manchester, England, on the 5th and 6th September 1998.

Section 2 describes some of our previous experiences with Inhabited TV and the particular problems which we encountered in combining multiuser virtual environments with television-style broadcasting. Section 3 describes and contrasts related work. Section 4 gives further information about the structure and nature of the OOTW event, and also describes the first two techniques which we used to enhance the broadcast potential of the show: appropriate world design; and managed participant constraints. Section 5 describes the third technique: a virtual camera control interface tailored for real-time performance. Section 6 contains our evaluation of the camera control interface and the event's (television-based) approach to mixing and directing, based on a social scientific field study conducted on-site. Finally, section 7 contains our conclusions and describes areas for future work.

[1]School of Computer Science and IT, Email: {cmg, sdb, imt}@cs.nott.ac.uk
[2]Centre for User-Oriented IT-Design (CID), Email: bowers@nada.kth.se
[3]Email: graham.walker@bt-sys.bt.co.uk
[4]Email: john@illumin.co.uk

2 EXPERIENCES WITH INHABITED TELEVISION

The authors have been involved in three previous experiments in inhabited television [2]:

- “MASSIVE” [3], a virtual poetry performance, occurring simultaneously in a physical theatre and in a virtual world. This was staged by the University of Nottingham, in collaboration with VR artist Sean Varney and poets from Chocolate Art, as part of Nottingham’s NOWninety6 arts festival.

- “The Mirror” [18], six public online virtual worlds run along-side the UK BBC television series “The Net”. This was created by BT and Illuminations, in association with the BBC and Sony. The broadcast in this case was based on edited recorded footage.

- “Heaven & Hell - Live”, a live 1-hour television broadcast on the UK’s Channel 4 from inside a public virtual world. This was also created by BT and Illuminations, in association with Sony.

In each event there were four types of participants (or “levels of participation”):

- passive viewers, either watching broadcast television or watching a single projected view in a theatre (in the case of MASSIVE);

- online inhabitants, embodied within the virtual worlds, and able to interact and take part;

- professional performers or correspondents, also embodied in the virtual worlds, providing content and coordination; and

- the production team, including camera operators, director and technical support, who worked behind the scenes to make each event happen.

These experiments raised a number of significant issues with the creation of inhabited television. The issues relevant to this paper are described below.

- It is typically very difficult to achieve precise and coordinated movement within a multiuser virtual environment. In a normal television show the positions and movements of the participants are tightly controlled, to support its presentation and timing. In a virtual environment movement is typically comparatively clumsy and uncontrolled. So it was hard for inhabitants (untrained members of the public) and even for performers to be in the right place at the right time.

- There is a huge difference in natural pace between virtual environments (as they exist now) and television. Television is typically tightly scripted and directed to give an intense viewing experience. General-purpose multiuser virtual environments, on the other hand, tend towards a much more sedate pace of interaction. This is especially true when interaction is text-based, as was the case in The Mirror and Heaven & Hell – Live (here we are ignoring specialized quick-reaction game environments, e.g. first-person shooting games or virtual racing, because we wish to explore a broader range of genres and richer forms of social interaction).

- There were significant problems with camera control and navigation. In particular, Heaven & Hell – Live and The Mirror used the normal user navigation interface to control the virtual cameras; this was a “body-centered” or flying vehicle control metaphor [19]. This meant that the virtual cameras had difficulty keeping up with the action and got lost on occasion, even giving shots of empty space. They also had trouble coping with occlusion in the scene – it was extremely difficult to adjust the shot to avoid, for example, a passing inhabitant, while maintaining the shot.

These problems combined to create a final broadcast which was often incoherent, and sometimes incomprehensible to watch as a normal viewer. These problems are demonstrated most clearly in Heaven & Hell – Live because it combines a truly live broadcast with a relatively complex and densely populated world.

As described in this paper, we have sought to alleviate these problems by using three techniques: appropriate world design (section 4.1); managed movement of inhabitants and performers (section 4.2); and a new performance-oriented virtual camera control interface (section 5).

3 RELATED WORK

Multiuser virtual environments, and film and television each have their own distinct bodies of literature. However, in this paper we are focusing on the boundary between the two: shaping and enhancing a live video output from a shared 3D audio-graphical environment.

In this event we adopted non-automated approaches to camera control and mixing, relying instead on established human skills and patterns of working, i.e. a director and camera operators. In this way we could explore how these activities were influenced by the change of technology and medium, rather than simply recreating and automating existing techniques which may not be appropriate for 3D virtual worlds.

This may be contrasted with attempts to automate the whole presentation process, including camera work, direction and mixing. For example, He, Cohen and Salesin [12] created the “Virtual Cinematographer”, also used in realtime multiuser virtual worlds. However, in their case, camera placement and cutting are fully automated. Their virtual environment also uses a constrained set of high-level actions (e.g. “go to the bar”) whereas our worlds are based largely on free navigation and interaction. This approach might be applicable in some situations, although it would need to extract likely semantics from more primitive activities (e.g. free navigation).

As our understanding of Inhabited TV matures it will become possible to integrate automation in a principled and informed way. This will allow us to draw on work such as the following. Drucker et al. [7] defined a procedural interface for expressing shots and integrating other techniques, which was then extended in [8] to use constraint-satisfaction, including path planning. Seligmann and Feiner [9] address automated placement of cameras and composition of shots for conveying information in semantically specified situations. They do not address live performance issues, although they indicate that their methods could be combined with interactive manipulation.

All of these approaches might be used to enhance or (perhaps partially) automate camera operation for Inhabited TV. However the realtime and the performance nature of the medium would have be carefully considered (an area which is not addressed in these papers).

Some of these, and others (notably Karp and Feiner [14], and He et al. [12]), have considered how cinematic principles (e.g. [1]) can be applied in computer graphics. In this work we are exploring how inhabited television and 3D graphical environments may actually differ from established experience with the physical world.

At a more conceptual level, this work can be compared to interactive digital artworks and installations, which combine computer graphics with aspects of live performance (see, for example, [17]). These same pieces might be made available through an inhabited television medium, although with an increased concern for the experiences of non-interacting (traditional TV) viewers.

4 OUT OF THIS WORLD

The Out Of This World event may be regarded as a follow on to Heaven & Hell – Live, taking into account the issues highlighted in section 2 (and issues relating to other aspects of the Inhabited TV). Like Heaven & Hell – Live, Out Of This World was a live game show staged in a multiuser virtual world. The game show was “broadcast” live into a theatre space, where it was projected onto a large screen which can be seen in figure 3.

The show featured a competition between two teams (the “aliens” and the “robots”), to escape a doomed space station (figure 5). Each team was led by a professional performer who was wearing an immersive VR headset and was tracked (head and both hands) using electromagnetic trackers. The team leaders are visible at the sides of the main projection screen in figure 3. Each team comprised four members of the public (drawn from those entering the theatre) seated at networked PCs in the back-stage production area, as in figure 4. In “real” Inhabited TV these people would be at home. These inhabitants could navigate on the ground plane using a standard PC joystick, and took part in four collaborative games. The show was fronted by a virtual host who appeared in the world as a live video texture on a large virtual screen (visible in figure 6). The inhabitants and performers all wore microphone-headsets and were able to talk to one another using packetised network audio (c.f. IP telephony). The host also had a microphone, and audio spot effects were triggered automatically at various points in the games. The multiuser virtual environment used was MASSIVE-2 [11].

To create the live audio/video feed into the theatre we made use of standard outside broadcast (field production) methods. Part of the back-stage area, including the director, is shown in figure 1. The video-related components of this are shown schematically in figure 2. There were four virtual cameras, each with its own operator. These fed into a video mixing desk (with input, preview and transmission monitors), together with a direct view of the host and a video tape (VT) player for additional linking footage. An experienced television director was responsible for directing and video mixing. An additional technician operated the VT, and audio was mixed separately by an audio engineer through the house PA. All of these people were on a wired talk-back circuit to allow them to coordinate their actions.

In this way we sought to exploit established methods of working and human skills (e.g. of the director and the camera operators) within a new context – 3D virtual worlds.

In the show’s finale the theatre audience were able to “save” one of the losing inhabitants using an interaction system inspired by Loren Carpenter’s CINEMATRIX Interactive Entertainment System [6], waving colored pieces of card to vote for their favorite player.

There were a total of four rehearsals and four performances over a period of three days, each lasting approximately 40 minutes.

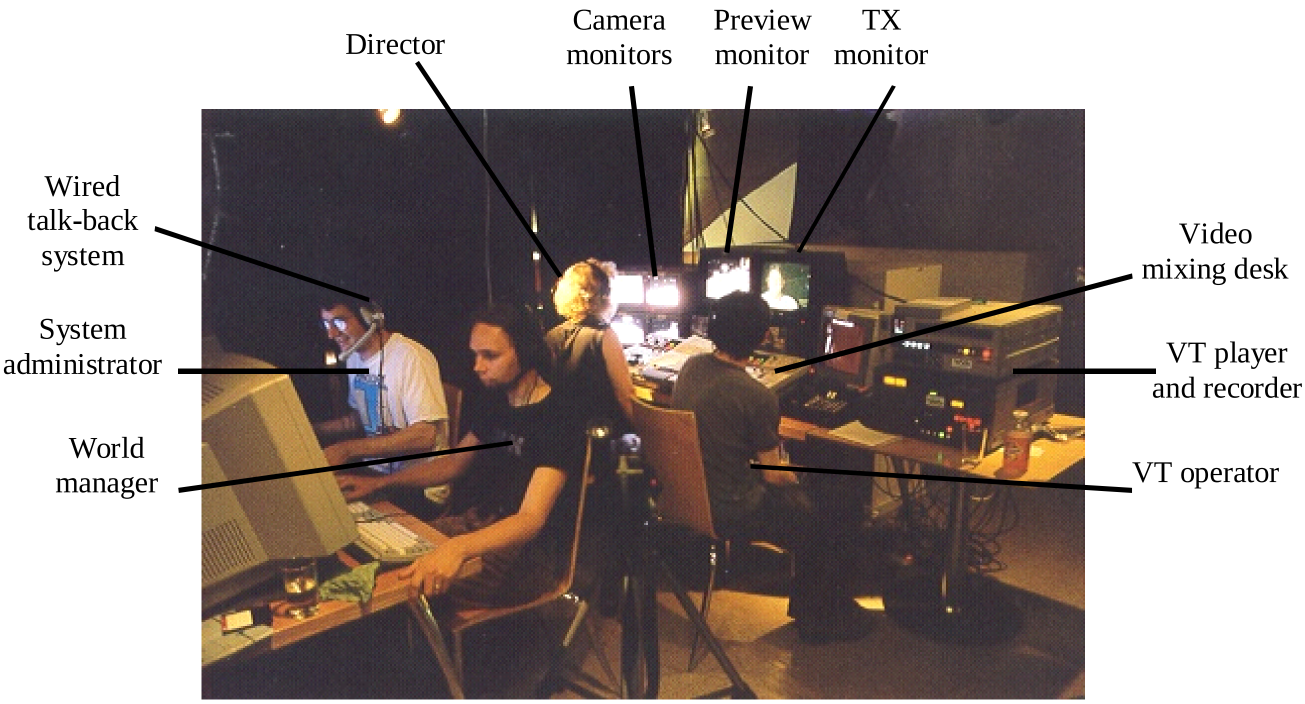

Figure 1: The director, VT operator and world manager at work, during a show.

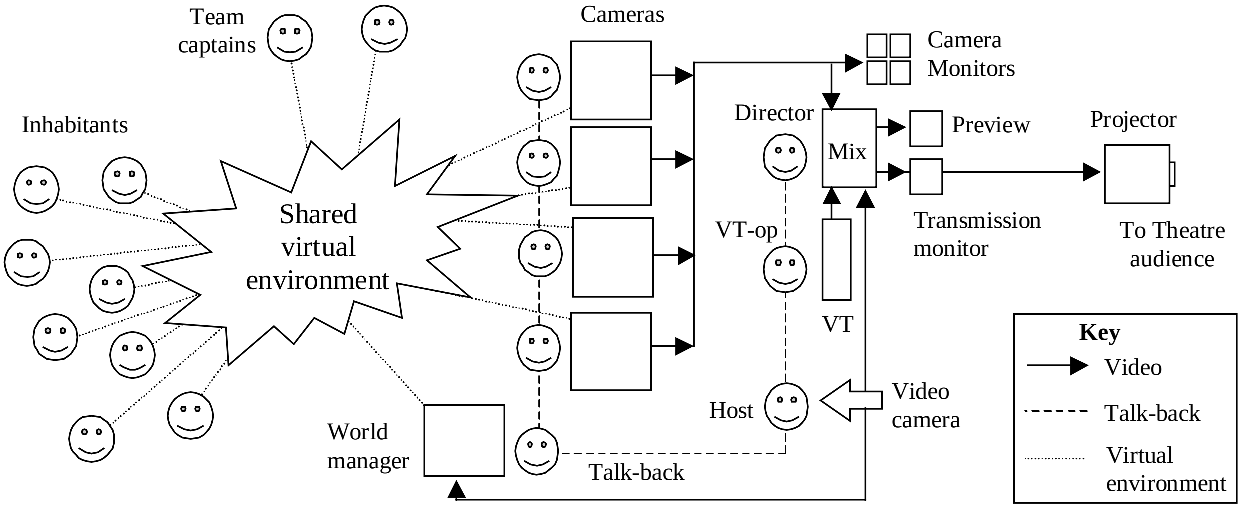

Figure 2: Video production infrastructure for OOTW.

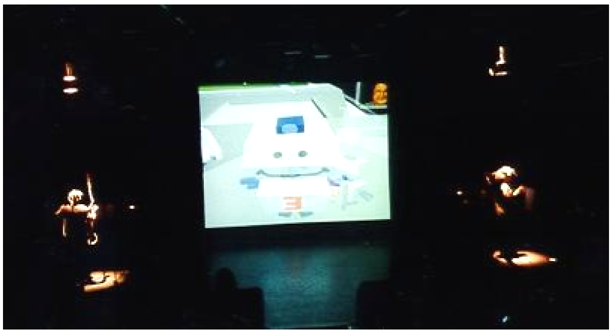

Figure 3: The front of the theatre.

Figure 4: The inhabitants, "at home", in the production area.

4.1 World Design Principles

In creating the world and games for OOTW we sought to enhance legibility, presentation and the effectiveness of the virtual cameras. To these ends we employed the following principles and guidelines:

- use simple representations, and simplify games and concepts wherever possible;
- exploit the system's realtime audio capabilities for interaction and effects;
- use an open world design with few visual barriers and no external walls (facilitating overview and context shots, such as figure 6) and use significant virtual separations to establish the different playing arenas (visible in figure 5);
- place most of the action at ground level; and
- use game interaction methods which rely on proximity, so that participants will be close to their effects in the environment (e.g. in one game – shown in figure 9 – inhabitants have to run up to "space frogs" to herd them towards their team captain; in another game the inhabitants all have to stand in a "jet car" and collectively guide it through a race course), rather than selection at a distance or shooting, for example.

In more constrained applications or genres some or all of these may not be possible or appropriate. We anticipate that, as the technology and common concepts of inhabited television evolve, many of these heuristics could be relaxed, or compensated for by additional automated assistance.

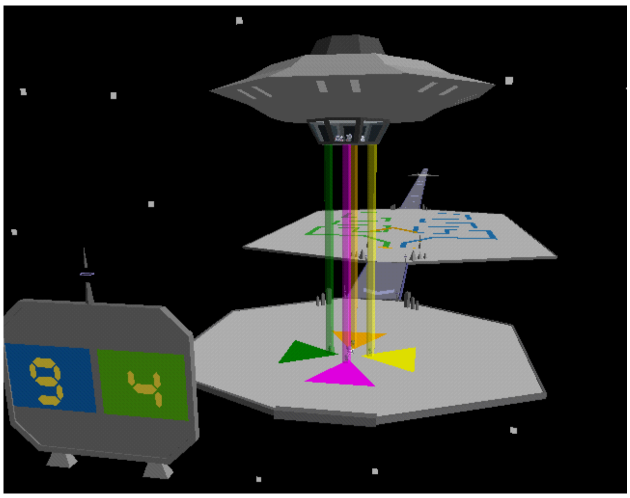

Figure 5: The virtual world, looking back past the final arena.

Figure 5 shows the scale and linear form of the virtual world. The camera is looking back past the host/scoreboard in the final arena (with the spaceship ready to leave), towards the race arena, with the previous (quiz) arena just visible in the distance. Figure 6 shows an aerial view of the quiz game (the third game arena). The host is visible as a video texture, overseeing the event. When the teams answer questions correctly their leader is raised up through the hoops by the world manager. Figure 7 shows the fish game in action (the second game arena). The inhabitants are collaboratively guiding their team leaders through the arena, to allow them to hook "space fish" down from the sky.

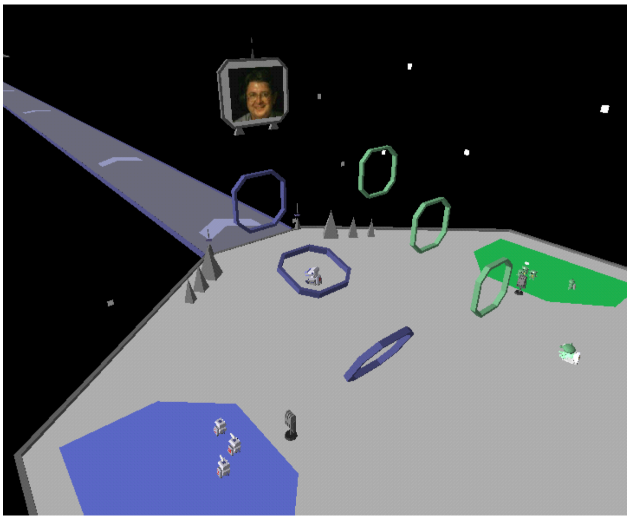

Figure 6: The quiz arena.

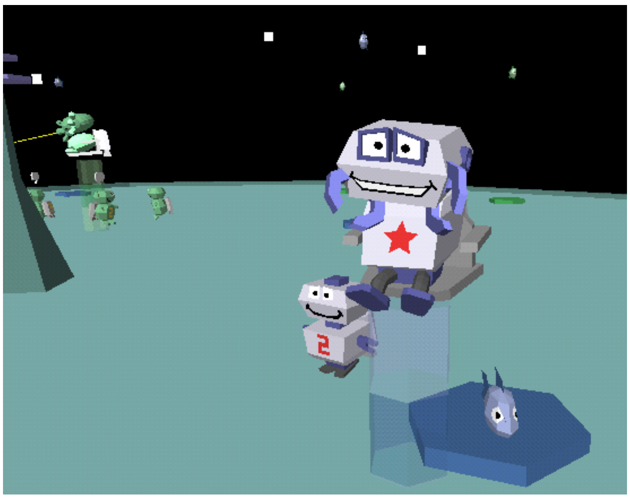

Figure 7: Inhabitants guide their team leaders in the fish game.

4.2 Participant Constraints

One of the key innovations of OOTW was the use of an event management and control interface which allowed us (amongst other things) to apply a positional constraint to each participant (inhabitant, performer or host). This constraint could be time-varying, allowing us to move participants from one part of the game world to another, for example creating virtual "travellators" between the different game arenas (visible in figures 5 and 6). These constraints could also vary in size, allowing us to "root" participants to the spot during interview segments or release them to explore a whole game arena (but not the whole world) during the actual games.

Specifically, each constraint consisted of a bounding box restricting movement in space, plus a forward (look) vector and permitted angular deviation. Constraints were defined in pairs, and linearly interpolated between, at a rate specified in each constraint-pair. The positional component and forward vector of each constraint were normally extracted automatically from annotations in the world definition files.

New constraints could be specified for every "phase" of the game, with the whole OOTW event comprising more than fifty phases, including contingencies and alternatives. The world manager (who was listening to directions on talk-back) was responsible for triggering phase changes from a list in the management interface. This management interface then informed each participant's client program of any new constraints, and triggered any required reconfigurations of objects within the virtual world (e.g. resetting game objects).

This allowed us to ensure that all participants kept up with the game – by dragging them if necessary – and were in the right place at the right time. In this way we addressed (at least partially) two of the issues from section 2:

- we could enforce coordinated movements and "crowd" control within the virtual environment, avoiding inhabitants getting lost or running out of control; and
- we could increase the pace of the event by enforcing time limits and by rapidly shuttling participants to new locations or elements of the show.

In addition, the use of constrained positioning meant that potential shots could be established in advance and the director and virtual camera operators could plan for them accordingly (e.g. establishing tailored camera positions and settings for different stages of the show, in relation to these pre-set constraints). For example, figure 8 shows the alien team being constrained to line up at the end of the final race. The inhabitants were visually alerted when their movement was constrained to avert frustration and confusion.

This same management interface also allowed the world manager to change the audio level of each participant within the virtual world, either to make them louder or quieter, or to virtually gag them, if necessary.

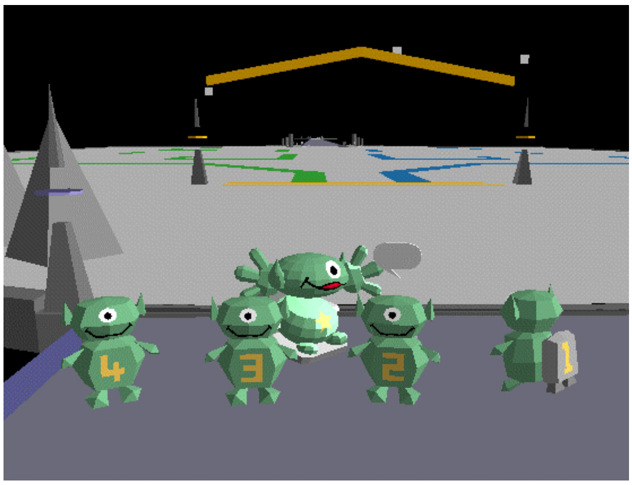

Figure 8: Participant constraints in action – lining up the aliens.

5 REAL-TIME CAMERA CONTROL INTERFACE

The previous section has described how appropriate world design plus constrained and managed participant movement were used to enhance the imageability of Out Of This World. In addition we created a specialized real-time camera control interface for the four camera operators. The characteristics of this interface reflect the real-time performance-oriented control requirements which derive from a live broadcast. In this section we consider the capabilities, construction and operation of this interface.

Figure 9 shows a screen image from one of the virtual cameras, showing both the normal flying vehicle-style controls (the arrow buttons at the bottom of the screen) and the custom camera interface (the dialog box at the right of the screen). The 3D window is sent to the mixing desk using the system's normal video output capabilities (SG Octane, Impact and O2 machines, with standard video options, were used for the cameras). The 3D view is taken from the "space frogs" game: robot number one, in the background, has just chased a space frog towards the robot team leader, ready for it to be impaled on the team leader's spiky hat.

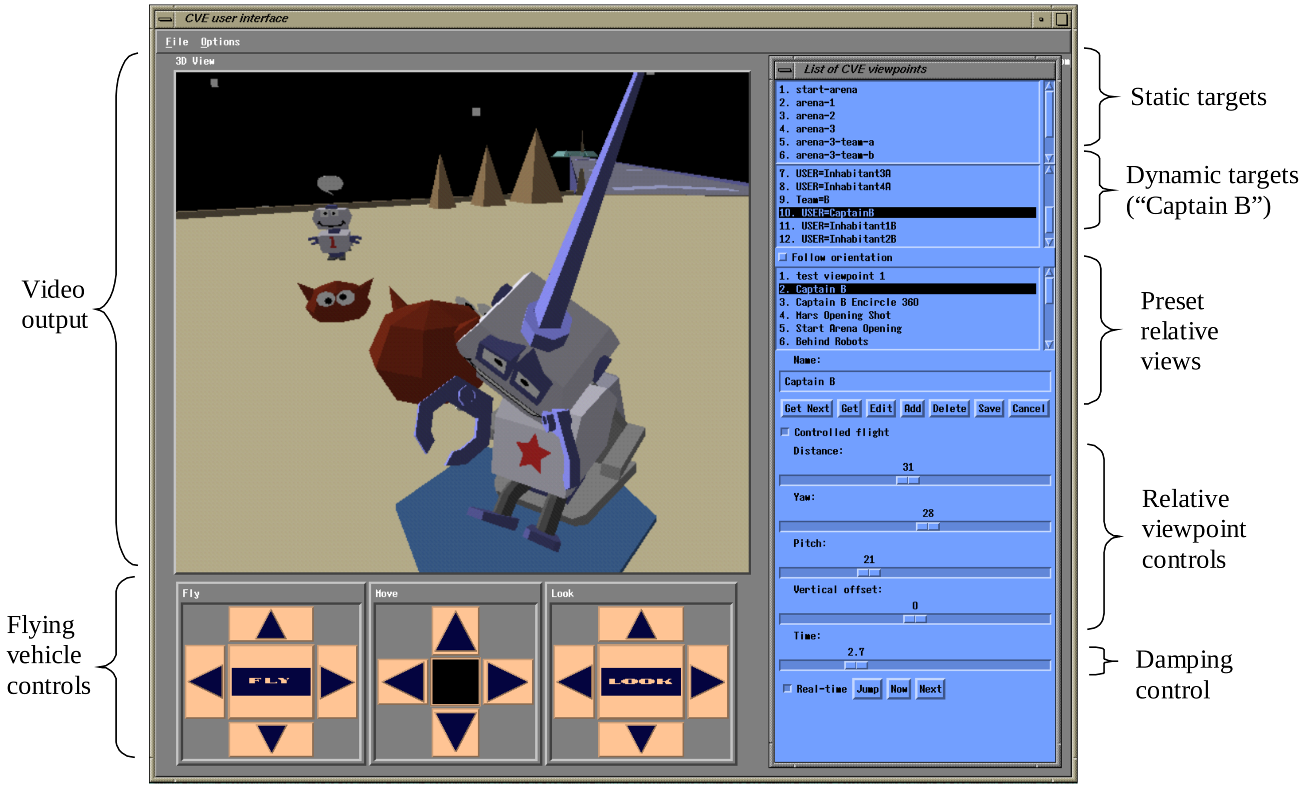

Figure 9: Camera interface for Out Of This World.

5.1 Flying Vehicle Control

The basic flying vehicle controls allow the camera to be moved in five degrees of freedom; the horizon was constrained to be horizontal at all times due to previous problems users had found with reorienting the viewpoint when rolling around the look vector. The main method of use is to control the camera's forwards/backwards velocity and left/right pan using the central black drag-box (which uses a cubic mapping from mouse displacement to vehicle velocity). Ware and Osborne [19] found this kind of view-relative navigational style (in their case controlled with a 3D tracker) to be good for "movie making", with its smooth control and perceived flexibility. The flying vehicle controls can also be used in concert with the object-centered controls described in the next section.

As noted in section 2, using only this type of interface in Heaven & Hell – Live produced a number of problems, such as failing to keep up with the action, getting lost, and difficulties coping with occlusion in the scene. Section 6 describes how it was used in Out Of This World.

5.2 Object-Centered Control

The flying vehicle metaphor controls both the position and orientation of the camera. The alternative object-centered controls allow the camera operator to specify independently the target of the camera (the look-at point as in [4]), and the relative viewing attitude and distance of the camera. This interface also provides additional temporal controls, as described below.

5.2.1 Look-at Point Control

The look-at point can be controlled in three ways:

- Vehicle-based free navigation, as described above, except that it is the target point rather than the camera itself which is being moved. This offers direct control over positioning in 3D space.
- Jump to preset fixed points in the virtual environment (derived from reference point artefacts included in the world

specification), for example, game arenas and player starting positions.

- Continuously track a single participant or a group of participants (e.g. team A or team B). The camera process continually monitors the activities of objects within the virtual world and adjusts the target as they move.

Note that the last two options are semantic choices expressed in terms of named objects or regions within the space. Their inclusion allows the camera operator to jump to key locations or track participants or activities irrespective of their speed (both of which were problems in previous experiments). In this respect the system introduces elements of constrained and automated camera control, as in [7] and [9], except that it is in the context of real-time control for performance.

The selected target can also optionally determine the reference orientation for the camera in the XZ plane (the ground plane). For example, when used with the tracking option it allows the camera to adopt first-person perspectives (as the participant sees the world), locked facing shots (keeping the participant full-face at all times) and a range of other possibilities.

5.2.2 Relative Viewing Control

In the object-centered control mode the position of the camera is not specified directly. Instead, it is specified relative to the position (and optionally the orientation) of the camera's current target. The relative position is specified primarily in terms of spherical polar coordinates, an approach which we first made use of in [3]. Specifically, the camera operator uses independent sliders to specify:

- the yaw of the camera relative to the reference orientation;
- the elevation of the camera;
- the distance of the camera from the target (including the option of going beyond the target to show an out of the eyes view for a participant); and
- a vertical offset relative to the target, to allow the operator to create views over or under the reference point, or to compensate for movements away from the ground plane.

A logarithmic control mapping has been used for the distance from the target, allowing controlled transitions between close and distant shots as proposed in [15]. These degrees of control are illustrated in figure 10.

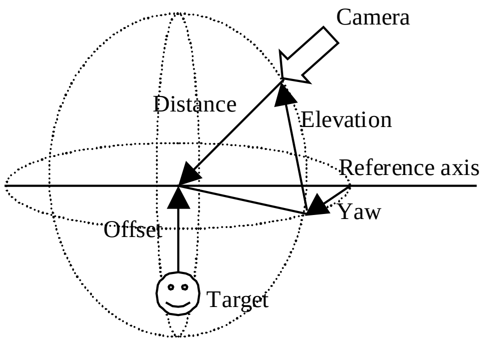

Figure 10: Relative viewing controls.

We suggest that this simple slider-based control over rotation is appropriate because it has been observed to be effective for single axis manipulation [5], and in our system we cannot use controls which overlap the 3D view because we cannot allow the mouse cursor to enter the graphical view (since it would be seen on the final broadcast).

These controls together allow the camera operator significant (but not complete) control over the framing of the target within the shot. They also (we hope) make it relatively difficult to create incomprehensible or empty shots.

5.2.3 Temporal Control

The final significant aspect of this control interface is its support for various forms of temporal control. In the current system the target point is controlled directly and interactively. However the relative viewing parameters can be controlled in three different ways, described below (and which may be compared to the automated mix-down capabilities of a modern digital recording studio).

- Real-time: the camera moves as each slider is moved, subject to a controllable damping coefficient which allows the operator to trade responsiveness against smoothness of movement.
- "Just-in-time": the camera operator can temporarily "disconnect" the sliders from the camera, make coordinated adjustments to all aspects of the relative view control and then trigger those changes as a single atomic operation. Again, the damping control determines the rate at which the camera interpolates between old and new settings.
- Pre-programmed: the camera operator can define complete sets of viewpoint parameters and store them for later recall. These preset viewpoints can then be selected by the operator as a single action, triggering an interpolation to that new viewpoint (with the rate determined by the damping value in the viewpoint being recalled).

The just-in-time and pre-programmed modes also allow the operator to specify a sequence of camera movements, to be executed one after another. Alternatively the operator can jump directly to the new viewpoint, or immediately start a new interpolation.

6 EVALUATION AND REFLECTION

As part of our evaluative reflection on OOTW a social scientific field study (of the type widely applied in the CSCW community, e.g. [13]) was conducted, with an ethnographer on-site during the performances of the show and the production and preparation work surrounding them. Here, we report on those aspects of the study concerned with the camera control software which forms a main focus of this paper.

6.1 Camera Use and Roles

The director of OOTW (RB) allocated the four camera operators to different basic tasks. One operator was to follow the activities of one team, another was to follow the other. A third was to get overall views of the environment ("geography shots"). A fourth was given a freer rein, instructed to seek out "relationships of interest, like the hand held camera would do". This division of labor maps well to the different forms of camera control provided for in the design we have discussed. The relationship operator, for example, might be expected to utilize the 5 DOF flying vehicle mode, while the team-oriented operators might be expected to use the option to track team-members and leaders. Finally, the operator seeking geography shots could adopt positions based on the static targets defined in the world definitions.

Interestingly, RB divided up the labor of camera operation in this fashion because this would be a standard division of labor for real television analogues of OOTW, a division of labor which was appropriately embodied in the software, and *not* because she saw these as the constraints built into the camera control technology. Clearly, future systems should also consider distinctive support for these same roles: geography or context shots; main characters or teams; and interest or relationship shots. Of course, this list cannot claim to be comprehensive.

During the later performances and as the operators' experience in virtual camera control increased, all operators were observed using the less constrained control modes (e.g. the 5DOF flying vehicle) more commonly. A manually controlled shot in pursuit of a team-leader might even be preferred over automatically targeting them. At least two reasons can be suggested for this. First, manual control can give the right amount of "camera shake" (as the target slips to one side or even momentarily out of shot) to convey a sense of frenetic activity. Second, manual control can enable the camera operator to follow the action in more flexible ways. For example, shots that were not directly supported by the camera control software could be achieved manually.

However, the controlled and preprogrammed shots we have discussed remained useful, especially in three contexts. First, when there were more scripted and recognizably repeatable moments in the action. For example, just after each game, the reactions of team leaders were asked for by the show's presenter (as in figure 8). Clearly, having standard methods to obtain such predictable shots is appropriate. Second, tracking shots were needed to follow rapid transitions, such as the movements between the different game arenas. Third, as RB said in a briefing meeting for the camera operators: "If it all goes pear-shaped [badly wrong], I can always say go back to your terms of reference". Having a known responsibility for each operator and a series of standard shots associated with it provided the camera operators with "escape routes" and a "safety net" in times of trouble, which wouldn't be available from higher DOF navigation vehicles alone.

Perhaps this should not come as a surprise, since it echoes common ideals of user-interface design: that the interface should be accessible to relatively novice users; that it should make common jobs easy; and that it should allow experienced users to exploit further levels of flexibility and customization.

6.2 Camera Coordination and Embodiment

One of RB's "worst nightmares" was that all cameras at a specific moment would head for the same shot and that she would have "nothing to cut to". For example, something especially remarkable at a particular location might catch the eye of all operators. In real television, the physical embodiment of the cameras and their operators militates against this to some extent: if a camera physically moves in a certain direction to get a shot, it is often clear to others what is going on. Conventional cameras also have facilities for operators to check out the views from other cameras without having to release their own current shot. Furthermore, operators typically have visual access to a "TX monitor" (showing the transmitted shot) commonly placed on the studio floor. In this way, operators can have an awareness of what each other are doing and check what is being transmitted. This can all help find an optimal shot or angle and avoid this problem.

However, in OOTW, it was decided not to graphically represent the cameras in the game worlds for other, good reasons (e.g. to avoid distracting the participants or occluding their views). Also, the only way the operators could see what others were capturing was by physically looking over to their workstations. Additionally, the operators had no TX monitor. This led to a number of occasions where RB had a sub-optimal selection of shots available to her, especially during the less scripted, more free-form action components of the show. Some clear enhancements suggest themselves, for example, access to a TX monitor and to the viewpoints of other cameras. Note that this is not concerned so much with an individual operator's camera control application or interface, but with how a set of cameras work together in the cooperative activity of inhabited television direction and camera work.

Also, the disembodiment of the cameras meant that performers and inhabitants did not know when and in what way they were being filmed, and so could not produce their speech and actions with reference to the cameras. In a future event it might be possible to use "subjective" views, to allow performers and inhabitants (and also camera operators and the director) to see the cameras, while removing them from the final broadcast output.

Finally, in OOTW both the director and crew only had visual access to the world through what the cameras were picking up, so there was no way to "have a quick look round" (e.g. to find who is speaking). This suggests the provision of extra views to camera crew and director may be necessary, so that they can explore the environment for their own purposes without this being made available for transmission.

6.3 Camera Interface and Video Output

In OOTW the 3D graphical views of the camera operators were directly output as the final broadcast material. This created five significant limitations:

- It was not possible to allow the camera operators to see the other cameras embodied without them also appearing on the final broadcast.
- The camera operators had very limited screen space available for additional interface controls or feedback mechanisms (such as an embedded TX monitor).
- The camera operators had to use a primarily 3D graphical view of the world, rather than, for example, a schematic or plan view of the environment, which might have been more effective in some situations.
- The camera operators had to avoid getting their mouse in the main graphical view because it would have appeared in the broadcast. Consequently they could not use image space controls such as Gleicher and Witkin's [10] "through-the-lens" method of incremental camera control.
- The image quality and frame-rate of the final broadcast was exactly that of the camera machine currently being broadcast. This meant that camera machines were a critical limitation in the broadcast quality, and motivated against additional computational or rendering complexity (and, therefore, capability) in the camera interfaces.

We have recently prototyped a "software mixing" solution which avoids this problem, and opens up opportunities to directly address the limitations described above.

In this approach, the final broadcast is generated by a single (typically high-performance) rendering machine, which runs the same multiuser virtual environment software. The viewpoint of this rendering machine is controlled directly by the currently selected camera, i.e. the camera machines output sequences of viewpoints rather than actual video. Note that the director also requires an additional machine to provide camera and preview monitors, and has to cut between cameras using the computer interface, rather than a traditional video mixer.

6.4 Manual Control

We believe that the design strategy to facilitate manual control rather than to automate shots and their transitions was appropriate. In particular, for research purposes, preferring enhanced manual control at this stage of research allows the users, as part of their actual interaction with the technology, to express in an occasioned and natural way the benefits and difficulties they experience in practice. It is these expressions which we have summarized in this section and are able to use as a source of future requirements.

Furthermore, limiting the degree of automation has allowed the director we have worked with to experiment with different ways of directing inhabited television. She can vary the pace of cutting, give different sets of instructions to the camera crew, get into the midst of the action or view it from the side, and so forth. Indeed, she consciously experimented with different direction styles during the course of the four OOTW shows, deciding to make fewer cuts but getting the cameras closer in for the last two shows. Again, it seems appropriate to design software at this stage which will enable such experimentation from television personnel, permitting them to address issues about the nature of inhabited television from their own professional viewpoint (e.g. how should one cut between virtual cameras?) rather than mandating an answer through excessive automation in software design.

6.5 Live Performance Issues

From section 5.2.3 it is apparent that temporal control over viewpoint in live performance settings can be more complicated than in non-realtime or non-performance applications. For example, in both computer animation and interactive exploration compound camera movements can be made incrementally (e.g. using a version of the real-time interface); it does not matter if the camera passes through intermediate "incomplete" states. However in a live performance, every placement of the camera must be acceptable in itself, since it may be part of the broadcast. To address this, the just-in-time mode of the camera control interface allowed the camera operator to compose complex camera movements to be executed as a single action. Similarly, the pre-programmed control interface allowed the camera operator to recall complete viewpoints which were established in advance.

Both of these facilities are motivated by the live performance nature of the interface's use. In a general sense, they address the extremely time-limited nature of live performance, by allowing the operator to exploit work done during less constrained periods of time. The pre-programmed interface allows the operator to use time in rehearsal or training to establish viewpoints for time-critical use in the performance. Similarly, the just-in-time mode allows the operator to accumulate unseen work in order to define a complex viewpoint transition.

We suggest that any live performance of significant richness will need to draw upon a significant body of previously established resources (e.g. script, training, understanding, virtual world definitions, geometry, computer programs). This work points to the potential for directly supporting this within the production and broadcast infrastructure.

7 CONCLUSIONS AND FUTURE WORK

From audience feedback and our own reviewing of the show video tapes we find that the broadcast output was legible and coherent (unlike some of our previous experiences) and the pace of the event was much closer to television than to normal interactive virtual environments. We ascribe this success to the following factors: relatively simple world geometry, structures and games; management controls to constrain the movements of participants; a specialized realtime performance camera interface; and increasing experience with events of this kind and their appropriate visual form (for example, reducing the cutting rate).

Having demonstrated the ability to create pace and legibility in an established and cliched game show format we now intend to develop this work in the areas of richer content, more complex narrative forms, more expressive avatars and new forms of interaction. We plan to stage our next major experimental event late-1999. The final sections describe some of the areas in which this work could be taken forward.

Camera Mixing

In OOTW we performed the final video mix in the analogue video domain using standard video mixers. However, using a software mixing approach (as described in section 6.3) we can use a single "broadcast" machine to render all of the final broadcast. The main advantages of this approach are that: the broadcast machine can be more powerful than the cameras, giving a better quality output (e.g. higher framerate); automated or semi-automated mixing can be integrated and used in parts of the broadcast (see also the next section); and the camera operators can have more specialized interfaces, for example including direct manipulation within their view (e.g. using the methods of [10]). In future experiments we plan to make use of this approach.

Automation

In Out Of This World, we did not attempt to define camera shots or transitions automatically. However there are many situations in which partial or full automation of camera positioning and/or mixing would be advantageous. We are particularly interested in using techniques for automated viewpoint control and/or mixing to reduce production costs, with a view to supporting long running, possibly permanently available, inhabited television experiences. Alternatively, automation might be used some of the time or for some aspects, supplemented by human expertise and aesthetic guidance.

Mix Your Own

With a traditional broadcast the individual viewer has no control over what they see. If the image from each camera was individually broadcast then the viewer would be able to "mix their own" final program by switching between the individual camera channels. Alternatively, if the virtual world data itself is broadcast then the viewer could choose between the director's mix, their own mix, or controlling their own independent viewpoint. With a suitable network back-channel, they could also make the transition to become an inhabitant, becoming embodied, active and visible within the event itself.

For Those Watching at Home…

The staging of OOTW fell short of "real" Inhabited TV in one important respect: neither the viewers nor the inhabitants were actually at home, in front of their TV/set-top box. In our next set of experiments we will return to domestic viewers and inhabitants, as in The Mirror and Heaven and Hell – Live. For viewers, the removal of the social context of the theatre space is likely to make them more demanding of the broadcast in and of itself. For the inhabitants, use from home raises major issues of heterogeneity and reliability of hardware and software, and of trust, accountability, control and governance for the individual inhabitants.

Audio

This paper has focused on the visual aspects of inhabited television. However, audio also plays a critical role. In film and television there is typcally a very rich and complex relationship between sound and vision, and it will be a significant challenge to support some of this richness in an Inhabited TV system. By analogy with virtual cameras, we also plan to explore approaches to virtual microphones and appropriate sound design, production and presentation for inhabited television.

Acknowledgements

We acknowledge the support of the European Community for the eRENA project under the ESPRIT IV Intelligent Information Interfaces (I3) programme and also the EPSRC for the Multimedia Networking for Inhabited Television project under the Multimedia Networking Applications programme. Additional support for Out of this World was provided by the UK's FACT and North West Arts.

We would like to thank all those who helped to design, create and stage Out Of This World – especially Mike Craven, Jason Morphett, CJ, Gareth, Jules and Rena – and all those who took part as inhabitants and viewers.

References

[1] D. Arijon. *Grammar of the Film Language.* Communication Arts Books, Hastings House, New York, 1976.

[2] Steve Benford, Chris Greenhalgh, Chris Brown, Graham Walker, Tim Reagan, Paul Rea, Jason Morphett and John Wyver. Experiments in Inhabited TV. *CHI'98 Late Breaking Results (Conference Summary)*, pp. 289-290. ACM, 18-23 April 1998. ISBN 1-58113-028-7.

[3] Benford, S.D., Greenhalgh, C.M., Snowdon, D.N., and Bullock, A.N.. Staging a Poetry Performance in a Collaborative Virtual Environment. *Proceedings ECSCW'97*, Lancaster, UK. Kluwer, 1997.

[4] Blinn, J.. Where am I? What am I looking at? *IEEE Computer Graphics and Applications*, July 1988, pp. 76-81.

[5] Michael Chen, S. Joy Mountford and Abigail Sellen. A Study in Interactive 3-D Rotation Using 2-D Control Devices. *Computer Graphics*, 22(4), August 1988, pp. 121-129.

[6] CINEMATRIX, Interactive Entertainment System, http://www.cinematrix.com/ (verified April 28, 1999)

[7] Steven M. Drucker, Tinsley A. Galyean and David Zeltzer. CINEMA: A System for Procedural Camera Movements. *Proc. 1992 Symposium on Interactive 3D Graphics*, Cambridge MA: ACM Press, March 29-April 1, 1992, pp. 67-70.

[8] Steven M. Drucker and David Zeltzer. Intelligent Camera Control in a Virtual Environment. *Proceedings Graphics Interface '94*, pp. 190-199, 1994.

[9] Dorée Duncan Seligmann and Steven Feiner. Automated Generation of Intent-Based 3D Illustrations. *Computer Graphics (SIGGRAPH 91 Conference Proceedings)*, 25(4), July 1991.

[10] Michael Gleicher and Andrew Witkin. Through-the-Lens Camera Control. *Computer Graphics*, 26(2), July 1992, pp. 331-340.

[11] Greenhalgh, C. M. and Benford, S. D.. Supporting Rich And Dynamic Communication In Large Scale Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments*, Vol. 8, No. 1, February 1999, pp. 14-35, MIT Press.

[12] Li-wei He, Michael F. Cohen and David H. Salesin. The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing. *SIGGRAPH 96 Conference Proceedings*, Annual Conference Series, pp. 217-224, Addison Wesley, August 1996.

[13] John Hughes, Val King, Tom Rodden, and Hans Anderson. Moving Out from the Control Room: Ethnography in System Design. *Proc. CSCW'94*, October 22-26 1994, Chapel Hill, NC, USA, pp. 429-439, ACM Press.

[14] Peter Karp and Steven Feiner. Issues in the Automated Generation of Animated Presentations. *Proceedings Graphics Interface '90*, 14-18 May 1990, Halifax, Nova Scotia, pp. 38-48.

[15] Jock D. Mackinlay, Stuart K. Card and George G. Robertson. Rapid Controlled Movement Through a Virtual 3D Workspace. *Computer Graphics*, 24(4), August 1990, pp. 171-176.

[16] Cary B. Philips, Normal I. Badler and John Granieri. Automatic Viewing Control for 3D Direct Manipulation. *1992 Symposium on Interactive 3D Graphics*, ACM Press, March 29-April 1, 1992, pp. 71-74.

[17] Ritter, D. The Intersection of Art and Interactivity. *Ars Electronica Festival 96*, Vienna: Springer, pp. 274-285, 1996.

[18] Walker, G.R.. The Mirror – reflections on Inhabited Television. *BT Technology Journal*, 16(1), pp. 29-38, 1997.

[19] Colin Ware and Steven Osborne. Exploration and Virtual Camera Control on Virtual Three Dimensional Environments. *Proceedings of the 1990 Symposium on Interactive 3D Graphics*, Special Issue of Computer Graphics, Vol. 24, pp. 173-183, 1990.