THE QUEEN'S AWARDS
FOR ENTERPRISE
2001

BBC

R&D White Paper

WHP 128

December 2005

3D in Content Creation and Post-Production

O. Grau

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 128

3D in Content Creation and Post-Production

Oliver Grau,

Abstract

This paper is on the use of 3D data for content creation in TV- and film productions. Computer graphics methods are used for a wide range of applications in content creation, like the planning of productions, for pre-visualisation on set and when 3D graphics appear in the final programme. The degree to which 3D graphics are involved ranges from their use in the planning phase of a production without any graphical elements in the final programme, to films that are entirely computer generated. For most types of programme it is in between and typically involves a mixture of virtual and real scene content. Well known examples are special effects in movies, which insert virtual objects or characters into real camera footage or virtual studios in TV- productions that insert a presenter or actor into a completely virtual environment. Since the application of computer graphics is still very expensive usually only those components that can not be filmed will be generated virtually.

The final high-quality graphics for film- and many TV- productions are mostly generated off-line in a post-production phase. For on-set pre-visualisation and many TV-productions, that are either broadcast live or have a lower budget than feature film productions, the graphics are generated in real-time. The budget and the decision whether the image generation has to be in real-time constrain applicable techniques. For real-time TV productions usually only simple 2D scene composition can be used. In virtual studio systems for example, a camera image of the presenter or actor is taken in a (real) studio and then keyed out and overlayed onto a synthesised (virtual) image.

How convincing the composited result looks in terms of realism depends on the quality of the optical integration of the virtual and real scene components. In a full optical integration virtual and real objects are optically interacting with each other, that means they are occluding each other and cast shadows. With a 2D composition these optical interactions can only be established in a very limited way. For sophisticated optical interactions more 3D knowledge of the real (and virtual) scene is needed and the scene composition is done in the 3D domain.

The technical challenges of those productions that integrate virtual and real scene components are the subject of this paper.

This document was originally published in O. Schreer, P. Kauff, T. Sikora (eds.), *"3D Videocommunication: Algorithms, concepts and real-time systems in human centred communication"*, John Wiley & Sons Ltd., July 2005,
ISBN: 0-470-02271-X

1

Applications of 3D Videocommunication

1.3 3D in Content Creation and Post-Production

Oliver Grau, BBC R&D, United Kingdom

1.3.1 Introduction

This chapter is dedicated to the use of 3D data for content creation in TV- and film productions. Computer graphics methods are used for a wide range of applications in content creation, like the planning of productions, for pre-visualisation on set and when 3D graphics appear in the final programme. The degree to which 3D graphics are involved ranges from their use in the planning phase of a production without any graphical elements in the final programme, to films that are entirely computer generated. For most types of programme it is in between and typically involves a mixture of virtual and real scene content. Well known examples are special effects in movies, which insert virtual objects or characters into real camera footage or virtual studios in TV- productions that insert a presenter or actor into a completely virtual environment. Since the application of computer graphics is still very expensive usually only those components that can not be filmed will be generated virtually.

The final high-quality graphics for film- and many TV- productions are mostly generated off-line in a post-production phase. For on-set pre-visualisation and many TV-productions, that are either broadcast live or have a lower budget than feature film productions, the graphics are generated in real-time. The budget and the decision whether the image generation has to be in real-time constrain applicable techniques. For real-time TV productions usually only simple 2D scene composition can be used. In virtual studio systems for example, a camera image of the presenter or actor is taken in a (real) studio and then keyed out and overlayed onto a synthesised (virtual) image.

How convincing the composited result looks in terms of realism depends on the quality of the optical integration of the virtual and real scene components. In a full optical integration virtual and real objects are optically interacting with each other, that means they are occluding each other and cast shadows. With a 2D composition these optical interactions can only be established in a very limited way. For sophisticated optical interactions more 3D knowledge of the real (and virtual) scene is needed and the scene composition is done in the 3D domain.

The technical challenges of those productions that integrate virtual and real scene components are the subject of this chapter.

In order to achieve photo-realistic images of virtual scenes it is necessary to consider several optical phenomena. We here consider a *scene* as a set of observable objects and a set of relevant physical phenomena, that can explain the visual appearance of the scene.

Object types of particular interest include:

A *body* as a coherent area in 3-space; often only opaque bodies are considered. For this case they can be represented equivalently as volumes or surface models. A *light source* sends light into the scene, so that bodies become visually apparent. A *real camera* is the sensor of the visual appearance of the real scene. Further a *virtual camera* is a model of a real camera that is used to generate images from a scene description, i.e. a virtual scene.

Both computer vision and computer graphics have developed models for these phenomena. For the creation of content that involves the integration of virtual content into real scenes (or vice versa) it is important that optical interactions between the integrated objects are harmonised.

A requirement for a believable integration is that both the virtual and the real scene are registered in the same co-ordinate space. In particular a *matching camera perspective* of both the real and virtual camera is very important for a convincing integration of real and virtual scene components. Depending on the distortions of the real camera, a more or less complex camera model has to be considered, i.e. at least camera pose and focal length and in some cases centre point shifts and radial distortions are used.

The most important *optical interactions* between scene objects are:

- *Occlusions*, which if not considered would destroy the realism of an integration.
- *Shadows* are important because they give a visual cue. Without shadows objects are appearing floating in space and look very unnatural.
- *Reflections* are important for shiny objects, but less often implemented because they can be avoided by careful selection of the scene elements.

Further to these phenomena there are many other physical effects that can be observed like refraction, transparency, particles (smoke, fog). These are modelled and implemented in computer graphics and also used in some productions. However, a harmonisation between them for real and virtual scenes is rarely considered, mainly because of a lack of analysis tools and models for real scenes.

Not an optical interaction, but an often used feature is a *view interpolation or extrapolation.* In this case the virtual and real cameras are not matched, instead the virtual camera moves independently. This is often a desired feature for special effects or for visualisation like in sports when a virtual camera flies over or around a scene. In order to be able to do this, more knowledge or data of the real scene must be available. Therefore, for a maximum of freedom, a 3D model of the real scene is required.

1.3.2 Current techniques used for integrating real and virtual scene content

The simplest integration of virtual and real content is a 2D/2D composition, as still in daily use in TV productions like the weather forecast or standard news graphics. Normally the actor or presenter will be separated from the studio background using chroma-keying. This requires a specially equipped studio. Since there is no use of any 3D data the optical interactions are limited to carefully chosen scene layering: that means the presenter is always in front of the synthetic background. Camera moves are not possible.

The main innovation in the 1990's virtual studios was to attach a real-time camera tracking system to a studio camera. There are several tracking systems commercially available. One example of a camera tracking system developed by BBC [Thomas et al. (1997)] is based on coded targets mounted on the studio ceiling, as depicted in Fig. 1.1. An auxiliary camera attached to the studio camera is mounted looking upwards. A real-time hardware system computes the accurate 3D position and orientation of the camera at 50 fps from those images. In addition zoom and focus can be retrieved through lens-attached sensors.

The measured parameters of the real camera are transfered to a virtual camera. That means the virtual camera parameters are cloned from the real camera, which is also called the master camera in this case. The virtual camera is implemented based on a sufficiently fast graphics hardware to generate an image of the virtual scene with parameters changing in real-time. The composition is usually done with an external studio keyer, that provides a chroma-key of the studio camera image and overlays it onto the virtual background. Since the virtual camera has no knowledge about the real scene it is not possible to implement other optical interactions than basic occlusions, although most keying systems provide the feature to pass the shadow of the studio floor onto the virtual image giving the impression of having a shadow cast from the actor onto the virtual background. This only works for a flat virtual floor.

In TV productions virtual studios have found several applications, but conventional sets are still predominantly used. One of the reasons for that is because virtual studios require a chroma-keying facility that has to be installed permanently to be cost effective. Moreover, it requires expensive equipment and skilled operators and designers for the virtual sets. Therefore, virtual studios are mainly used for series with very short turn-around times, that means several different sets are used in the same studio even on the same day, or if the content is already highly multi-media oriented and the programme benefits from the use of virtual techniques.

In the film industry the use of chroma-keying is a very common production technique, in particular for special effects. Techniques developed for virtual studios are used for pre-visualisation on set. For example the BBC camera tracking system has been used in movies like A.I.[Rosenthal et al. (2001)] and others to give the director a

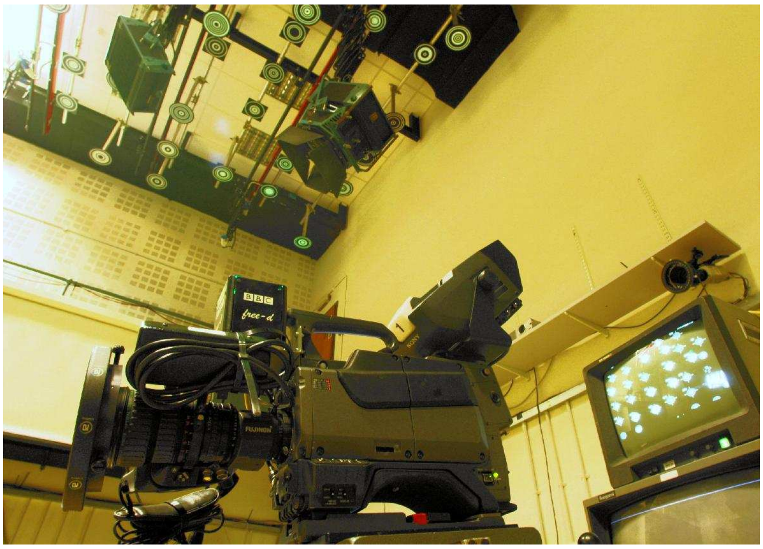

Figure 1.1: The BBC "Free-D" camera tracking system is based on circular bar-coded targets on the ceiling and an auxiliary camera looking upwards. Real-time hardware computes the position and orientation of the camera at a 50 Hz update rate.

pre-view of the camera framing with the action of actors and virtual backgrounds live on set. The final composition is done in post-production in a time consuming offline process that can take several days or even weeks to be finished. Therefore having the direct feedback available during the filming reduces the risk of a re-shoot of a scene. Unfortunately the actor is still working in a chroma-key environment and gets usually no direct feedback of the virtual scene.

A trend in recent TV productions is not to replace the entire environment with a virtual background. Instead only selected virtual scene components are integrated into a camera image. An example of such a production is depicted in fig. 1.2. The BBC programme is designed for children and is based on two teams. Each team designs its own small virtual robot and 'programs' it with a set of rules. The team's robots compete with each other in a virtual race run on the top of a large table. The robots are simulated in a games engine in real-time. A virtual image of the current scene is overlayed onto an image from a studio camera that is equipped with a tracking system. In order to cast a (virtual) shadow on the table, a model of the table in the form of a plane and the position of the studio key light is required. The shadow is then generated by the rendering software of the game engine.

An important part of the production is that children of both teams are present in the studio to 'support' their robot. A problem that occurs immediately is that the children need a feedback of the virtual robots, otherwise they can neither keep eye-contact nor react to the actions of the robots. Therefore an image of the virtual scene was projected onto the table in front of the children in the blanking phase of the studio camera, so that it is only visible to the children's eyes. The projection does not take into account the position of the viewers therefore the images are perspectively not absolutely correct, but they are sufficient for the visualisation purpose, since the children only have to see where the virtual robots are and what they are doing.

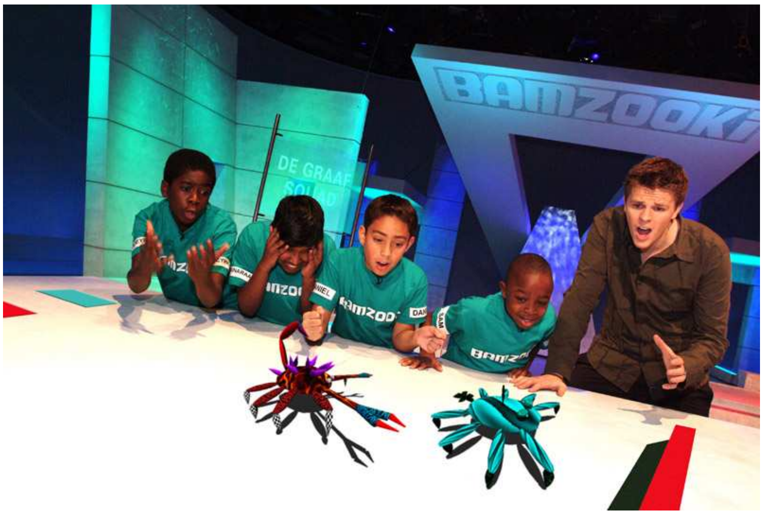

Figure 1.2 BAMZOOKi

Beyond the previous scenario many production situations ask for a bi-directional interface between the virtual and real world. For the previous given example that would mean that the children can not only see the virtual objects but would further be able to touch them or give them a spin. This feature can be achieved by tracking the actors movements and by detecting collisions of the real and virtual world. Further the position of actors can be used as a target for virtual characters to look at. Section 1.3.4 introduces a concept that is implementing a bi-directional interface.

Special effects found in feature film productions are normally not restricted to real-time and use more powerful rendering tools, like ray-tracers for image synthesis. Therefore, special effects can afford to establish full optical interactions, that means occlusions, shadow casting and receiving, plus reflections if needed. Control over the lighting is very important, some effects ask even for control over the virtual camera.

In order to be able to implement these sophisticated optical interactions the virtual and the real scene have to be represented in 3D. Once both are described as a 3D model the scene composition can be done with production tools like animation packages, providing a full optical integration.

How the creation of the 3D model of the real scene is done is very individual to each production and specific scene. Under many conditions the model is created interactively with an animation package, based on maps, measured dimensions of the set or based on images of the set or objects. If a higher level of detail or accuracy of the objects is required then an automatic modelling approach will be used.

The automated creation of 3D models of static objects makes use of active, sensor-based or passive, image-based methods. Active, sensor based 3D reconstruction systems, as described in chapter 5.1, use an active illumination technique and are usually more robust compared to most passive techniques. There are several products available using active techniques from companies such as lasers or structured light. For the generation of both static objects and also complete sets laser-scanning systems are predominantly used by the film industry for special effects.

Passive, image-based 3D reconstruction techniques use only camera images without any active illumination. An example of this class of techniques is stereo vision, as described in chapter 3.2. For the creation of 3D models of static objects or sets these methods are not much used in productions, mainly due to their restrictions in accuracy and robustness.

Concerning the estimation of the real lighting situation, it is common practice to take a 'light-probe', i.e. a picture with a white sphere in the scene. In an animation package a virtual white sphere is placed and virtual lights are placed interactively until the rendered image matches the image with the real sphere. In recent years this approach has been supplemented by image-based methods. In particular high-dynamic images of the scene, as described by Debevec (1998) can be used with many industrial rendering packages to re-create even complex lighting situations.

1.3.3 Generation of 3D models of dynamic scenes

3D models of dynamic objects like actors are usually generated with automated methods, since it requires a model update for every frame of a sequence. The same set of methods as discussed before for static objects can be considered, but only a few meet the requirements to capture the 3D shape and texture of an object in real-time.

Many active or sensor-based systems are not fast enough to capture at film or video frame rate. For this reason image-based methods are used or systems with active illumination based on cameras that can capture images of the light pattern and the texture. Such systems have been used for the real-time capture of faces. A structured light approach is commercially available from Eyetronics. In the movie the 'Matrix Reloaded' a stereo reconstruction was used to generate 'clones' of persons.

For the 3D-capture of complete objects, in particular of actors in a studio environment, the computation of the visual hull from 2D silhouettes has been shown to be quite robust and suited for many practical object classes [Cheung et al. (2003); Grau and Thomas (2002); Matusik et al. (2000)]. The approach is based on the shape from silhouette method as described in chapter 3.3, which is relatively simple to implement and can be computed in real-time in a specially equipped studio using chroma-keying or difference-keying techniques.

A disadvantage of the basic shape from silhouette algorithm is that no convex structures can be modelled. This problem was addressed by several extensions of the approach. The voxel colouring and shape carving technique [Kutulakos and Seitz (2000); Seitz and Dyer (1999)] makes use of the colour information, i.e. the differences between the generated model and the camera images. A more robust and often computationally faster approach can be achieved by combining shape-from-silhouette with stereo matching [Starck and Hilton (2003)]. The application of stereo requires usually a different camera setup than a visual hull computation (pairs of relatively close grouped cameras). Another extension of the basic shape from silhouette algorithm is to consider the temporal changes as in [Vedula et al. (2002)].

A different strategy, that fundamentally makes use of silhouette information is the incorporation of high-level generic models of human bodies [Carranza et al. (2003); Hilton et al. (1999); Weik et al. (2000)]. These methods initialise a generic human shape model and then update only motion parameters over time. They give a good appearance, but are limited to human object classes.

Three-dimensional models of the actors are used for the real-time on-set visualisation and off-line for post-production. In the first case a rough quality is sufficient since the purpose is to generate a pre-visualisation, but it has to be in real-time. In the latter case the aim is to get a 3D quality that can be used for special effects in final programme quality. In both cases the shape-from-silhouette or visual hull concept can be used. For the real-time version the technique has to be computationally very fast. Several fast methods for the visual hull computation have been proposed [e.g. Matsuyama et al. (2004); Matusik et al. (2001); Niem (1994)].

For use in the final programme the 3D models are usually computed off-line in post-production. Therefore more time consuming methods can be used in order to achieve more accurate and moreover temporally consistent 3D models. Most visual hull reconstruction algorithms compute a volumetric reconstruction in a first step and use an iso-surface generator, like the marching-cubes algorithm [Lorensen and Cline (1987)] to compute a 3D polygonal surface description in a second step. Possible artefacts in this reconstruction can be grouped into three categories:

- There is remaining volume due to occlusions
- Bad volume approximation due to low number of cameras
- Sampling problems

Fig. 1.3 illustrates the effects of these problems. The reconstruction in the left and middle was done using only six cameras and an octree-based data structure with hierarchical approach. The voxel resolution is 128x128x128.

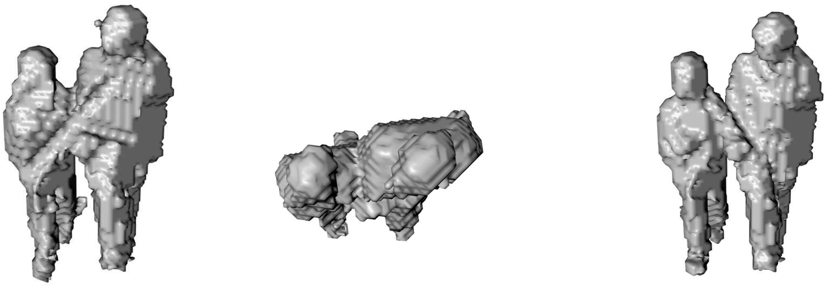

Figure 1.3: Result using 6 cameras and voxel-based 3D reconstruction (128x128x128) front view (left), top view (middle) and 12 cameras (right)

In left and middle views of Fig. 1.3 the problems due to occlusions can be seen: There are remaining bits on the feet and the separation of the two persons is quite poor. The approximation problems due to the low number of cameras are clearly visible in the top view (middle) of Fig. 1.3: In-particular the person on the right shows a 'blocky' shape.

Both problems can be overcome by using more cameras: By upgrading the studio system from 6 to 12 cameras, the shape quality is improved, as can be seen in Fig. 1.3 (right).

Unfortunately there are still very visible artefacts that are due to sampling problems. First the 3D reconstruction looks bigger than the actual objects. This effect is caused by the 2D box test that gives a 3D voxel model that is on the average 0.5 voxels bigger than the real object. Therefore by increasing the voxel resolution the model can be better approximated. On the other hand an increase in voxel resolution also increases the number of triangles of the resulting surface description.

Another visible problem in Fig. 1.3 (right) is the fact that the surface looks 'voxelised', meaning that the voxel structure is visible due to another sampling problem: The used voxels are binary, they indicate only whether that voxel is foreground or background. The marching cubes algorithm that is used to generate a surface description from the voxel representation is introducing quantisation noise for this kind of representation. One way to overcome this problem is using a line-based sampling method as described in Grau and Dearden (2003). In Grau (2004) we propose a method based on an octree-representation and super-sampling, that gives smooth 3D surfaces that can be used to generate video sequences. This approach reduces the sampling error that would otherwise be caused by a conventional volumetric reconstruction and the use of the marching-cubes algorithm for the generation of a surface model. The new approach extends the accuracy of the volumetric shape reconstruction by super-sampling without increasing the number of triangles in the 3D model: The leaf nodes of an octree are further subdivided and the value of the original node is replaced by a counter of the number of sub-nodes that are found as belonging to the object. This value is than used in a standard marching cubes algorithm to compute a smoother 3D surface. Fig. 1.4 (left) shows a resulting reconstruction using one super-sampling level. Further we apply Gaussian smoothing to the 3D models in order to suppress temporal artefacts that would be visible in a synthesised video sequence. Fig. 1.4 (right) shows an integration of the model into a 3D model of an entrance [1] with full optical interactions.

1.3.4 Implementation of a bi-directional interface between real and virtual scene

The interface between the real and the virtual scene components plays an important role during 3D content production. In particular the on-set visualisation of the virtual scene becomes an important issue for the actor, and the pre-visualisation of the composited scene for the director and operators on set. If the actor or presenter has

[1] The 3D model of the entrance was provided by Multimedia information processing group of Christian-Albrechts-University Kiel

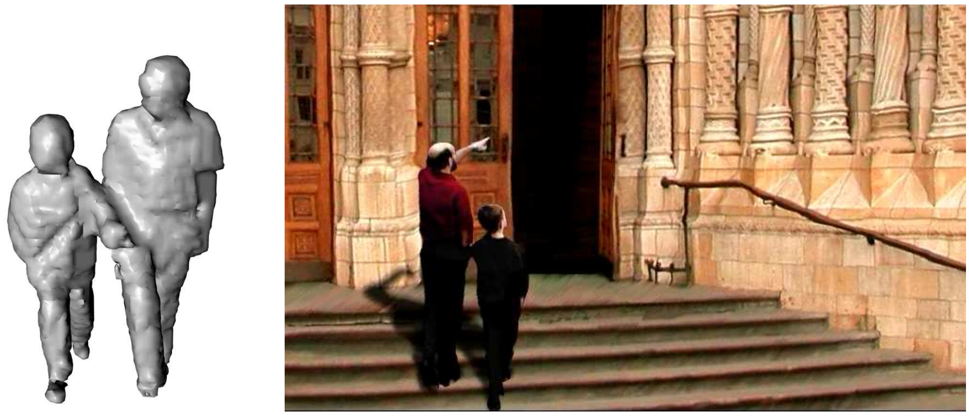

Figure 1.4: Result using voxel-based 3D reconstruction with super-sampling (left) and resulting frame of a video sequence with 3D model integrated into virtual background (right)

to interact with a virtual object, e.g. a virtual character, then without feedback there is often a noticeable difference between the direction in which the actor is actually looking and the position of the virtual character (the so-called eye-line problem). This can be quite disturbing in the final programme since human observers are very sensitive to wrong eye-lines.

On the other hand feedback from the real scene to the virtual scene is also very desirable in many production situations. It requires sensing of the real scene and can be used to generate precise event triggers or to create virtual objects that react to the real world.

An approach to provide the actor or presenter with a visual cue of objects in a virtual set is described in [Tzidon and et al. (1996)]. The system projects an outline of virtual objects onto the floor and walls. However, this method is restricted to show only the point of intersection of the virtual objects with the particular floor or wall. That means a virtual actor in the scene would only be visualised as footprints and the eye-line problem persists.

Systems that do provide the required functionality are projection-based VR (virtual reality) systems, like the CAVE [Cruz-Neira et al. (1992)] or the office of the future [Raskar et al. (1998)]. The main application of these systems is to provide an immersive, collaborative environment. Therefore, a CAVE system tracks the position of the viewer's head and computes an image for that particular view-point, which is then projected onto a large screen forming one wall of the environment. Several screens are usually used to provide a wide field-of-view, and some installations provide an all-round view using six projection screens that completely surround the viewer. Therefore it is possible to present objects that appear virtual in space, giving the viewer an immersive experience. The head position of the viewer is usually tracked using a special device, e.g. a helmet with an electro-magnetic transmitter, and if stereo projection is desired, a pair of shuttered glasses must be worn. Such devices

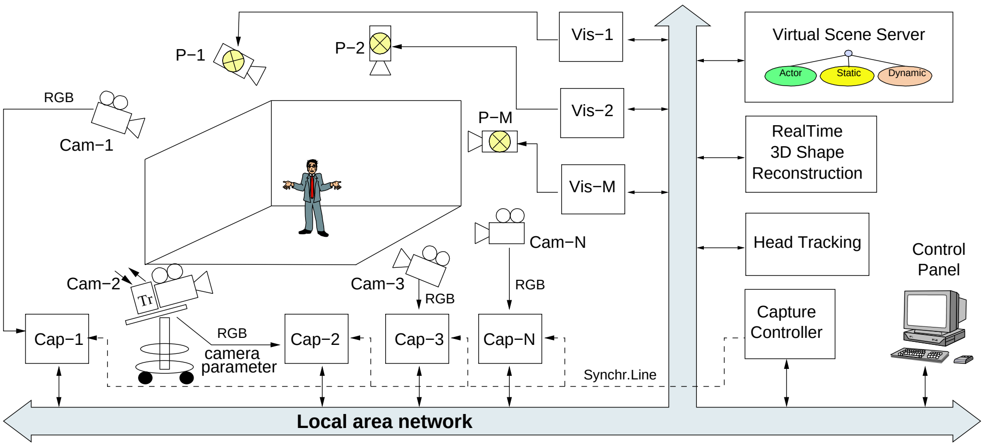

Figure 1.5: System overview. Each camera (Cam-1 .. Cam-N) is connected to a capturing server (Cap-1 .. Cap-N). Equally the data projectors (P-1 .. P-N) are driven by a projection server (Vis-1 .. Vis-N).

cannot be used in a production environment, because they would be visible in the final programme.

Although such collaborative projection-based VR systems would probably provide a good immersive feedback for an actor, they are not designed to create 3D models of the person. The main problem with the integration of a 3D capturing component into a CAVE-like system is that there should not be any interference between both subsystems. A concept of integrating 3D capture into a CAVE-like system was proposed in the blue-c system [Gross et al. (2003)]. This system uses back-projection onto a special screen that can be switched in three cycles between transparent and opaque modes. In the transparent mode images are taken in parallel from several cameras. In the opaque mode a projector projects an image for the left and right eye of the virtual scene in two separate cycles in order to give a stereo cue. The user has to wear shuttered glasses, that separate the images for both eyes. For the 3D reconstruction an approach similar to Matusik et al. (2000) based on difference keying is proposed.

The system discussed here was designed for the production of 3D content as part of the IST-ORIGAMI project [Grau and et al. (2003)]. Therefore, the quality of the final 3D content has the highest priority. Furthermore the system is intended to be used in a standard studio environment. Taking these requirements into account a view-dependent front-projection system was developed that requires less space than a rear projection system and can be fitted into most existing studios. Images are projected onto a special retro-reflective cloth that allows a robust chroma-key in conjunction with cameras equipped with a ring of monochrome (blue) LEDs.

The system is composed of a number of modular components, as depicted in fig. 1.5. The communication and data exchange between these components is predominantly based upon a local area network and standard IT components. This concept provides a cost-effective way of implementing a system which can be easily configured to suit particular requirements. For example the number of cameras can be varied depending on the available space in the studio and the specific production needs.

Each camera (Cam-1 .. Cam-N) is connected to a capturing server (Cap-1 .. Cap-N). The capturing servers are standard PCs, equipped with a frame grabber card. Each capturing server also has a RAID disc array. This allows the incoming video from the cameras to be saved to disc as uncompressed video (704 x 576 pixel, 24-bit colour resolution at 25 fps, progressive scan). This data is later processed in an off-line phase to produce high-quality 3D models for the final programme.

Usually the cameras are fixed and their parameters are determined with a calibration procedure. In addition, cameras equipped with a real-time tracking system can be used, as shown for Cam-2 in Fig. 1.5. The real-time tracker delivers exact position and orientation and the internal camera parameters. We are using our previously-developed 'Free-D' camera tracking system, as described in section 1.3.2 for our experiments.

In addition to their function as an image sequence recorder, the capturing servers provide several online services over the network. On request they can send the latest grabbed image and provide a chroma-keying service, allowing the other software components to request alpha masks. One of the components making use of this is the real-time 3D shape reconstruction, that synchronously requests alpha masks from all the capture servers, and uses these to compute a 3D model of the actor using fast visual hull computations as described in the previous section.

The 3D data is used by the head tracker and passed to the virtual scene server. The head position of the actor is used by the projection system to render a view-dependent image of the scene.

The virtual scene server provides a description of the virtual scene. This includes static scene elements, usually the virtual set, dynamic parts, for example any virtual characters involved in the scene, and the actor, provided by the 3D shape reconstruction module. The virtual scene server synchronises all scene updates and distributes the scene updates to it.

The visualisation servers (Vis-1 .. Vis-M) are standard PCs equipped with OpenGL accelerated graphics cards that render the scene and drive the data projectors (P-1, .. ,P-M). They access the latest scene data from the scene server and the actor's head position from the head tracking module. Each server uses this data, together with the calibrated position of the corresponding projector, to calculate and render the scene for the viewpoint of the actor. The generated image is then projected onto the retro-reflective cloth.

The control panel is the main interface to the system. It consists of several components:

The *camera remote control* allows the remote setting of camera parameters, like focus, zoom, aperture and so on. This is particularly important because one or more cameras may be hanging from the ceiling and hence will not be accessible.

The *capture control* is used to start and stop the capturing service on the servers.

The *animation control* is able to start pre-defined 3D animations or control them live.

The *3D preview* provides a preview of the production for the director and the camera operators. This allows the director to see a view of the scene from any position he chooses, allowing planning for camera shots. A textured 3D mesh of the actor is inserted in the virtual scene to give a preview of the final composited programme.

The control panel can be physically distributed over several workstations. This allows tasks to be delegated to different people or to be controlled from different locations. Moreover, the preview module can be instantiated several times, e.g. a view for the director, for a camera operator, for an animator and so on. The particular viewpoint of each preview can be set individually and is dynamic.

The capture controller synchronises control events and the actual capturing. Therefore all capture servers are connected with a hardware line (indicated as "Synchr.Line" in Fig. 1.5) that allows frame-accurate synchronisation of the capture.

A more detailed description of the studio set-up and the system components is given in Grau et al. (2004).

Head Tracking

The view-dependent actor feedback system requires information of the actor's head position. CAVE systems use mainly electro-magnetic or acoustical tracking devices that are attached to the auxiliary glasses (e.g. shutter or polarisation glasses), that are necessary for the stereo perception. For a production system this intrusion is not acceptable. Therefore a passive method is used. For this purpose a fast 3D template filter can be applied to the 3D volumetric representation generated by the visual hull reconstruction. The filter consists of two boxes, V_1 and V_2, as depicted in Fig. 1.6. The filter computes a match $M(\mathbf{P})$ for all positions $\mathbf{P} = (m, n, k)^T$ in a (discrete) volume:

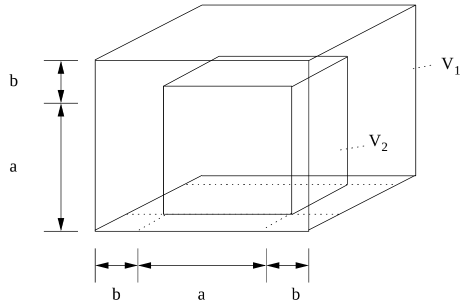

Figure 1.6 Template for head tracking

$$M(\mathbf{P}) = 2a_2(\mathbf{P}) - a_1(\mathbf{P}) \tag{1.1}$$

$$a1(\mathbf{P}) = Match(V_1(\mathbf{P})) \tag{1.2}$$

$$a2(\mathbf{P}) = Match(V_2(\mathbf{P})) \tag{1.3}$$

$$\mathbf{P}_{Head} = \mathbf{P} : M(\mathbf{P}) \rightarrow max. \tag{1.4}$$

The match functions in (1.2) and (1.3) count the numbers of voxels inside the test box $V_1(\mathbf{P})$ and $V_2(\mathbf{P})$. In practical use the method has shown itself to be quite robust.

Besides the use in a View-dependent projection system the actor's head position can also be used as a feedback source for virtual scenes. For example an autonomous (virtual) character can react to the actor's motion. Because the head position is precisely known the virtual actor could track the actor and keep its eye-line correct.

View-dependent rendering

The rendering engines in the visualisation server for the projectors request the latest head position from the head tracker and use this to calculate and render the projected image from the point of view of the actor. This is done by setting a viewing frustum with the origin at the actor's head position. In addition a homographic distortion (8 parametric perspective image transform) is added to the projection matrix of the virtual camera, taking into account the relative positions of the projection screen (wall) and the projector position.

The view-dependent rendering allows the actor, if required, to keep looking at the face of the virtual character as he walks around him. That means the virtual scene components appear in space and the actor is 'immersed'.

The renderer also receives any updates in the scene from the virtual scene server. If the virtual character was to move then the actor would see these movements. The combination of the scene updates and the viewpoint-dependent rendering thus allows complex interaction between the virtual and the real scene elements.

Fig. 1.7 shows the set-up during a demo production. The scene is taking place in the (virtual) entrance hall of a museum and includes a Pterosaur flying through the hall from one end to the other. The flight was visualised using four projectors and allowed the boy to precisely track the Pterosaur over 180 degrees. The boy's eye-line was perfect in all takes that were recorded.

Mask generation

A problem in a front-projection system is that there could be projected light falling onto an actor. One way to make this unwanted light invisible to the cameras is by operating the cameras in a shuttered mode, and projecting only during the time that the shutters are closed, as proposed in Tzidon and et al. (1996). One disadvantage of this approach is that the light intensity of the projectors must be very high in order for the actors to be able to see the projected image under normal studio lighting conditions. Furthermore, the need to shutter the cameras may force the use of higher levels of studio illumination. Also, there remains the potential problem that the actors may be dazzled when looking towards a projector. An alternative approach, proposed here, is to generate a mask from the captured 3D shape and overlay it onto the projected image. This removes the need for high-intensity projectors with a rapid temporal response, so that conventional data projectors may be used, and also solves

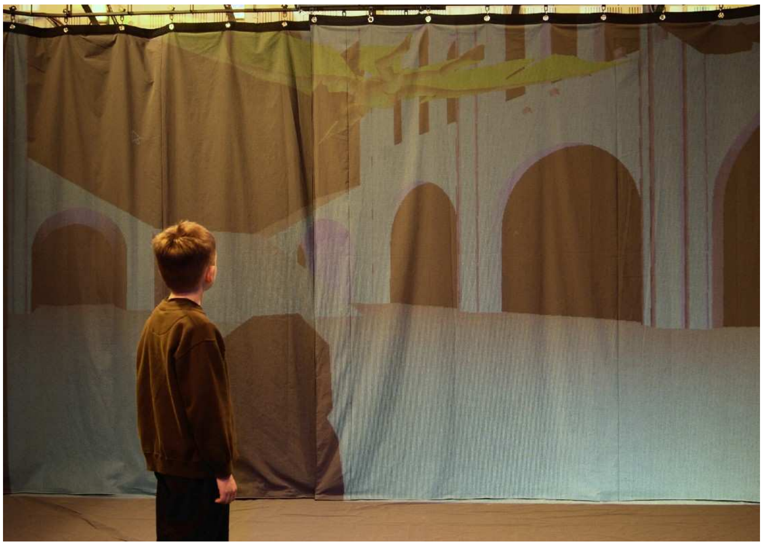

Figure 1.7: Flying pterosaur on the projection screen (in the top middle of the screen). The black area right in front of the actor is caused by a shadow mask.

the problem of dazzle. The mask can be seen in Fig. 1.7 as black area on the projection screen just infront of the actor.

Texturing

For the director or camera operator, a synthesised view is provided that gives a pre-visualisation of the final composited scene, i.e. the virtual and real scene elements. This renderer receives updates from the virtual scene server and grabs the latest 3D shape model of the actors. It can then generate a 3D representation of the scene and allows the director to view the scene from any position. This position can be dynamically updated to allow simulation of shots where the camera is moving.

In order to give a more realistic image the 3D shape model of the actor is textured with a view from one of the cameras. Therefore the renderer determines the studio camera that has the smallest angle to the virtual camera. The 3D shape model is stamped with the time-code of the alpha masks used to generate it, so the renderer requests the image from that time-code from the relevant capturing server, and uses it to texture the 3D shape model.

Collision detection

Collision detection is a computer graphics technique to detect whether two 3D objects are in contact or are intersecting each other. Here it is used to establish another source of feedback for the virtual scene. In particular the collision detection allows determination whether an actor is in contact with a virtual object.

In order to keep the latency low the collision detection should take place after the visual hull computation. Here the test was implemented in the volumetric data representation: an updated bounding box of the virtual object of interest is compared with the result of the visual hull reconstruction. An event is generated if the intersection increases beyond a certain threshold of voxels, which avoids firing events caused by noise in the volumetric reconstruction.

The output signals are sent to the event server, which re-broadcasts them to its other clients. The use of these signals to trigger external events is designed into the functionality of the clients (ie the studio renderers).

The collision detection has been used in tests together with the projection system. Because it is intuitive to use and the event triggers are very precise; the acceptance was very high.

1.3.5 Conclusions

3D data is increasingly used in content production, in particular in the movie industry for special effects, but also for TV productions. A driving force in this industry is the search for new effects or new kinds of programmes. Examples for that are movies like the Matrix with excessive special effects or completely new television programmes like the BBC BAMZOOKi programme.

On the other hand new production methods can significantly decrease the production costs. An example is the use for on-set visualisation that has already changed the production flow for special effects in the film industry. The discussed production system with a bi-directional interface between real and virtual world introduces a set of new tools for these kind of productions. It is worth mentioning that the degree of immersion in this system is not as high as in stereoscopic display systems, but it is a valuable tool to solve interaction problems between virtual and real actors, in particular the important eye-line problem.

For the future the increase in the use of 3D data and 3D scene composition will continue. To tap the full potential of cost savings, 3D planning and pre-visualisation tools will be further developed.

In addition the final programme content will benefit from new production methods and the use of 3D data, mainly for special effects, but also for new emerging techniques that use 3D delivery, like 3DTV and gaming platforms.

Bibliography

Carranza J, Theobalt C, Magnor M and Seidel HP 2003 Free-viewpoint video of human actors. *ACM Trans. on Computer Graphics.*

Cheung K, Baker S and Kanade T 2003 Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 77–84.

Cruz-Neira C, Sandin D, DeFanti T, Kenyon R and Hart J 1992 The cave: Audio visual experience automatic virtual environment. *Communications of the ACM* **35**(6), 65–72.

Debevec PE 1998 Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography *Proceedings of SIGGRAPH 98, Computer Graphics Proceedings, Annual Conference Series*, pp. 189–198, Orlando, USA.

Grau O 2004 3d sequence generation from multiple cameras *Proc. of IEEE, International workshop on multimedia signal processing shop on multimedia signal processing 2004*, Siena, Italy.

Grau O and Dearden A 2003 A fast and accurate method for 3d surface reconstruction from image silhouettes *Proc. of 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 395–404, London, UK.

Grau O and et al. 2003 New production tools for the planning and the on-set visualisation of virtual and real scenes *Conference Proc. of International Broadcasting Convention*, Amsterdam, NL.

Grau O and Thomas GA 2002 Use of image-based 3d modelling techniques in broadcast applications *2002 Tyrrhenian International Workshop on Digital Communications*, pp. 177–183, Capri, Italy.

Grau O, Pullen T and Thomas GA 2004 A combined studio production system for 3-d capturing of live action and immersive actor feedback. *IEEE Transactions on Circuits and Systems for Video Technology* **14**(3), 370–380.

Gross M, Wuermlin S, Naef M, Lamboray E, Spagno C, Kunz A, Koller-Meier E, Svoboda T, Gool LV, Lang S, Strehlke K, Moere AV and Staadt O 2003 blue-c: A spatially immersive display and 3d video portal for telepresence *Proc. of ACM SIGGRAPH 2003*, pp. 819–827, San Diego, USA.

Hilton A, Beresford D, Gentils T, Smith R and Sun W 1999 Virtual people: Capturing human models to populate virtual worlds. *IEEE Conf. on Computer Animation.*

Kutulakos K and Seitz S 2000 A theory of shape by shape carving. *Intl. Journal of Computer Vision* **38**(3), 197–216.

Lorensen WE and Cline HE 1987 Marching cubes: A high resolution 3d surface construction algorithm *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 163–169. ACM Press.

Matsuyama T, Wu X, Takai T and Wada T 2004 Real-time dynamic 3-d object shape reconstruction and high-fidelity texture mapping for 3-d video. *IEEE Transactions on Circuits and Systems for Video Technology* **14**(3), 357–369.

Matusik W, Buehler C and McMillan L 2001 Polyhedral visual hulls for real-time rendering *Proc. of 12th Eurographics Workshop on Rendering*, pp. pages 116–126.

Matusik W, Buehler C, Raskar R, Gortler SJ and McMillan L 2000 Image-based visual hulls In *Siggraph 2000, Computer Graphics Proceedings* (ed. Akeley K), pp. 369–374. ACM Press / ACM SIGGRAPH / Addison Wesley Longman.

Niem W 1994 Robust and fast modelling of 3d natural objects from multiple views *SPIE Proceedings, Image and Video Processing II*, vol. 2182, pp. 388–397, San Jose.

Raskar R, Welch G, Cutts M, Lake A, Stesin L and Fuchs H 1998 The office of the future: A unified approach to image-based modeling and spatially immersive displays. *Computer Graphics* **32**(Annual Conference Series), 179–188.

Rosenthal S, Griffin D and Sanders M 2001 Real-time compter graphics for on-set visualization: "a.i." and "the mummy returns" *Siggraph 2001, Sketches and Applications.*

Seitz S and Dyer C 1999 Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision* **35**(2), 151–173.

Starck J and Hilton A 2003 *Proc. of ICCV*, pp. 915–922.

Thomas GA, Jin J, Niblett T and Urquhart 1997 A versatile camera position measurement system for virtual reality tv production *Conference Proc. of International Broadcasting Convention*, pp. 284–289.

Tzidon and et al. 1996 Prompting guide for chroma keying. *United States Patent.*

Vedula S, Baker S and Kanade T 2002 Spatio-temporal view interpolation *Proceedings of the 13th ACM Eurographics Workshop on Rendering.*

Weik S, Wingbermuehle J and Niem W 2000 Automatic creation of flexible antropomorphic models for 3d videoconferencing. *Journal of Visualization and Computer Animation* **11**, 145–154.