THE QUEEN'S AWARDS FOR ENTERPRISE 2001

BBC

R&D White Paper

WHP 108

March 2005

A 3D production pipeline for special effects in TV and film

O. Grau

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 108

A 3D production pipeline for special effects in TV and film

O.Grau

Abstract

This contribution describes techniques for 3D modelling and sensing of dynamic studio scenes and different visualisation techniques for the production of special effects in an improved production pipeline. The main concept of this new production pipeline is the use of a shared 3D database that is used and edited through the different production stages. The benefit of this concept was demonstrated during different phases of an experimental production. In particular during the on-set phase, the system provides visualisation tools including a pre-view of the composed programme for the director or an immersive feedback system for the actor. The immersive feedback is implemented using view-dependent projection onto a special retro-reflective chroma-keying cloth and does not interfere with the shape capturing sub-system. Finally the generated dynamic 3D models were used in post-production to achieve new effects, like full optical integration of real actors with virtual backgrounds (shadow casting and reception) and fully free selectable camera viewpoint.

This document was originally published in Proc. of Mirage 2005 conference, Computer Vision/Computer Graphics Collaboration Techniques and Applications. March, 1-2 2005, INRIA Rocquencourt, France

Additional key words: 3D reconstruction, special effects, Film and TV production, mixed reality, on-set visualisation

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

A 3D production pipeline for special effects in TV and film

Oliver Grau

BBC Research & Development, Kingswood Warren, Tadworth, Surrey, UK.

ABSTRACT

This contribution describes techniques for 3D modelling and sensing of dynamic studio scenes and different visualisation techniques for the production of special effects in an improved production pipeline. The main concept of this new production pipeline is the use of a shared 3D database that is used and edited through the different production stages. The benefit of this concept was demonstrated during different phases of an experimental production. In particular during the on-set phase, the system provides visualisation tools including a pre-view of the composed programme for the director or an immersive feedback system for the actor. The immersive feedback is implemented using view-dependent projection onto a special retro-reflective chroma-keying cloth and does not interfere with the shape capturing sub-system. Finally the generated dynamic 3D models were used in post-production to achieve new effects, like full optical integration of real actors with virtual backgrounds (shadow casting and reception) and fully free selectable camera viewpoint.

Keywords: 3D reconstruction, special effects, Film and TV production, mixed reality, on-set visualisation

1. INTRODUCTION

The use of computer graphics (CG) in TV- and film productions is increasingly popular today. It allows the creation of special effects in feature movies, visualisation of facts in scientific programmes and many other applications. The production costs of programmes that involve CG are still relatively high. In particular they quite often bear high risks in the case something goes wrong in the production chain. A typical example would be a computer generated character that can not be integrated into real film footage, because the camera angle or the positions of (real) actors were wrong in the real camera images.

This paper describes a new production flow for the generation of special effects in TV- and film production. This production flow makes use of the 3D domain in all stages of the programme production. Further it describes: The 3D reconstruction techniques and visualisation techniques developed in the IST-ORIGAMI project.[1, 2]

1.1. Current practice

The production of any TV or film programmes that involve high-end CG special effects usually has three technical phases:

In the *planning phase* the conceptional ideas, usually laid down in a storyboard are transformed into a script together with a list of scenes and technical instructions how to obtain these.

In the *on-set phase* the filming takes place according to the script.

In the *post-production phase* the virtual content is integrated with the camera footage and the single scenes are edited to the final programme.

Previously in conventional productions (fig. 1) the 3D content was only used at the post-production stage and the virtual content was thus only visible after all studio recording was completed. On the real set the actors and the director and camera operators usually had only simple visual cues, like a mark on the floor where the virtual characters or objects should appear.

One significant problem that often occurs is that the camera framing is wrong at this stage. For example, this might happen if a virtual character is walking side by side with a real actor and because the camera operator cannot see the virtual one he might not leave enough space in the image frame. It is then difficult or impossible to integrate the virtual scene objects in post-production.

Further author information: E-mail: Oliver.Grau@rd.bbc.co.uk, web: http://www.bbc.co.uk/rd

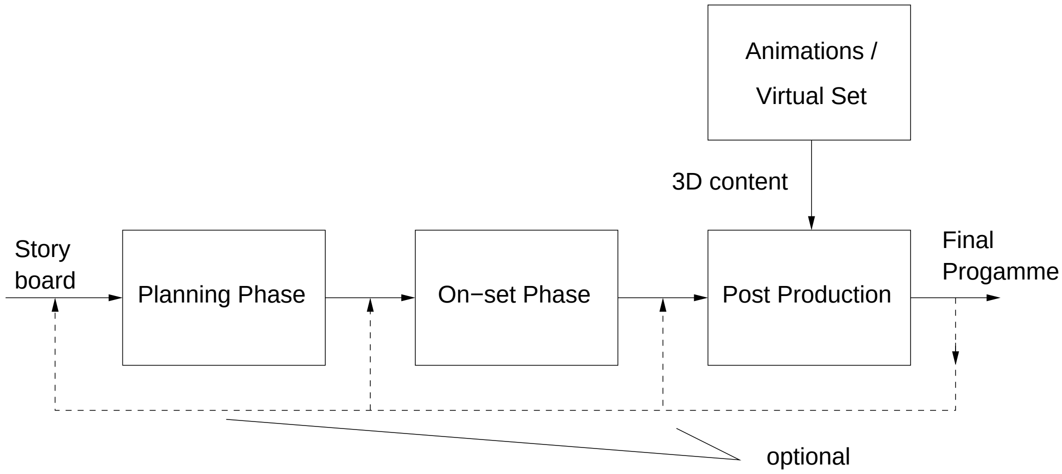

Figure 1. Conventional production work flow.

Another problem commonly found is the 'eye-line' problem. This means the actor is not looking exactly towards the virtual object, because it is not visible for him at this stage. This is very disturbing in the final programme, because human observers are very good in guessing the viewing direction of other people.

Bi-directional interactions between moving real and virtual characters are even harder to achieve, because they require an appropriate feedback of the virtual world to the actor and in some cases also feedback from the real to the virtual world.

In the worst case the filming on set has to be repeated, as indicated with the dotted path in fig. 1. Since this is very expensive the industry is aiming to avoid this.

1.2. Related work

In the EU-funded project IST-ORIGAMI[1,2] several techniques have been developed that allow the capture of 3D content for the production of special effects. This paper focuses on a review of the techniques used in the studio production phase.

Visualisation in the studio is an increasingly important problem. For the director, camera operators and actors there is a strong need to have an on-set pre-visualisation of the composited scene. This has been addressed in the past by using virtual studio techniques to get the composited scene on a studio monitor.[3] Unfortunately the actor still has the problem of looking into empty space. Hence synchronisation, particularly of the eye-lines with a moving virtual object is a major problem.

An approach to provide the actor or presenter with a visual cue of objects in a virtual set is described in.[4] 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.[5] 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.

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 sub-systems. A concept to integrate 3D capturing into a CAVE-like system was proposed in the blue-c system.[6] 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 is projecting time multiplexed images for both eyes in order to give a stereo cue. The user has to wear shutter glasses, that separate the images for both eyes. For the 3D reconstruction an approach similar to[7] based on difference keying is proposed.

Since shutter glasses would be visible in the final programme we proposed a less space occupying front projection system with a mono-scopic queue, i.e. no needs for auxiliary devices. Our system is based on a special retro-flective cloth,[8] that allows the actor to see the projected images, while the camera is equipped with a ring of blue LEDs. The light from the LEDs is reflected back to the camera and allows a robust chroma key. Section 3.1 gives more details on this technique.

For the generation of 3D models different techniques were developed by the ORIGAMI project[1]: For the generation of background models a structure-from-motion technique with plenoptic rendering was implemented. For single objects a level-set method was proposed. Alternatively manual creation with a 3D modelling package could be used.

The modelling of dynamic actors is described in the next section. It builds up on our work in[8, 9] based on a visual hull reconstruction from a multi-camera system.

In section 3 a new 3D processing pipeline is described that makes use of the 3D data from various authoring tools, including the dynamic models from the studio in a shared 3D database.

2. 3D MODELLING AND SENSING OF DYNAMIC OBJECTS

This section describes 3D modelling and sensing techniques for dynamic scenes in a studio with a multi-camera system that currently uses up to 12 cameras.

3D dynamic models are used at different qualities for on-set pre-visualisation and the final programme: For the on-set work a real-time visualisation is needed, but the visual quality does not have to be perfect, since it is only a means to allow the director to decide on positions etc. For the final programme the aim is to achieve the best visual quality for 3D models, but since it is an offline process more computation time can be used.

This section outlines briefly the methods implemented for 3D reconstruction, texture mapping and sensing the actor's head position and possible collisions with virtual objects.

2.1. Visual hull computation

The visual hull computation or shape-from-silhouette is a popular technique that gives relatively computationally fast and robust results.[10–12]

The approach requires a set of silhouette images from calibrated cameras. A silhouette image is a binary image where each pixel indicates whether this pixel belongs to the object or not. The silhouette information can be determined by any suitable segmentation process. In most cases this is done using chroma-keying or difference keying. In our studio system we are using chroma-keying with a special retro-reflective cloth and cameras equipped with a ring of blue LEDs. The light of the LEDs is reflected back by the special cloth and gives a saturated blue coloured background*.

The 3D shape reconstruction can be formulated as the intersection of generalised cones of the silhouette images. A generalised cone is the union of visual rays from all silhouette points of a particular image. This intersection gives only an approximation of the real object shape and is called the visual hull. In particular concavities cannot be modelled with this method, but for human beings this restriction can be tolerated.

For the computation of the visual hull many algorithms have been published, e.g.[10–13] These approaches solve the problem in a volumetric space representation. The most common of these representations is to subdivide a 3D box in euclidian 3-space into a set of voxels of discrete size. In order to save memory these are often represented as octrees[10, 11] or are run-length encoded.[12] An overview of volumetric reconstruction can be found in.[14]

For the use of the 3D reconstruction, e.g. in a computer graphic application, a surface description (usually a polygonal mesh) has to be generated. An algorithm often used is the marching cubes algorithm[15] which creates

*The cloth is commercially available under the name Chromatte™ from Reflecmedia.

an iso-surface of a volumetric data set. The marching cubes algorithm is quite robust and fast. Since the shape-from-silhouette methods mentioned compute binary voxels, the 3D surfaces generated from those using the marching cubes algorithm are very noisy. This noise is introduced due to the spatial discretisation of the volumetric representations.

We proposed two approaches to improve the computation of the visual hull of objects, by: a) a line-segment-based representation[16] and b) super-sampled octree representation.[9] The line-based approach is very fast and is therefore used in our real-time implementation.

For the off-line version we developed an octree-representation and a new 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 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 then used in a standard marching cubes algorithm to compute a smoother 3D surface. Further we are applying Gaussian smoothing to the 3D models in order to suppress temporal artefacts that would be visible in a synthesised video sequence. The result section shows an example of this approach.

2.2. Texture mapping

The texture mapping uses a directional mapping. For the real-time version the camera closest to the virtual camera is selected as a source for the texture map.

For the high quality rendering in post-production up to three cameras are used and blended.

2.3. Scene sensing

The calibrated multi-camera system allows image-based sensing of various parameters related to the actors in the studio:

The *head position* of actors is computed with a fast template matching filter in 3D space.[8] The filter works on the volumetric data. The resulting head position is used to drive the view-dependent projection, that is described in the next section.

A fast *collision detection* is also computed in the volumetric domain and checks whether the actor intersects with a virtual object. This can be used to precisely trigger pre-defined animations.

Both sensing methods are implemented in the studio system and are available as a service. Other modules can receive the information and it is possible to build up complex interactions between real actors and autonomous virtual actors.

3. A 3D PRODUCTION PIPELINE

One of the main weaknesses of the conventional production flow, as described earlier, is that the 3D data is first incorporated in the post-production phase. The work flow depicted in fig. 2 is based on a shared 3D data set.

In the planning phase the 3D data set consists of a (static) 3D model of the scenery. This model can be created by image-based methods, as developed by the ORIGAMI project[2] or manually with a 3D modelling package. The director is defining the action and the camera framing for each shot in this phase. The positions of actors can be marked by 3D place holders (generic human model) or simulated by an avatar animation package as proposed in.[1]

Since a 3D model of the set is already present at this stage the 3D animators can create a first draft of the proposed 3D animation with the virtual characters, or place (static) virtual objects depending on the programme. The resulting 3D animations can be rough at this stage, i.e. they do not provide final quality, because they might be updated in the on-set or post-production phase.

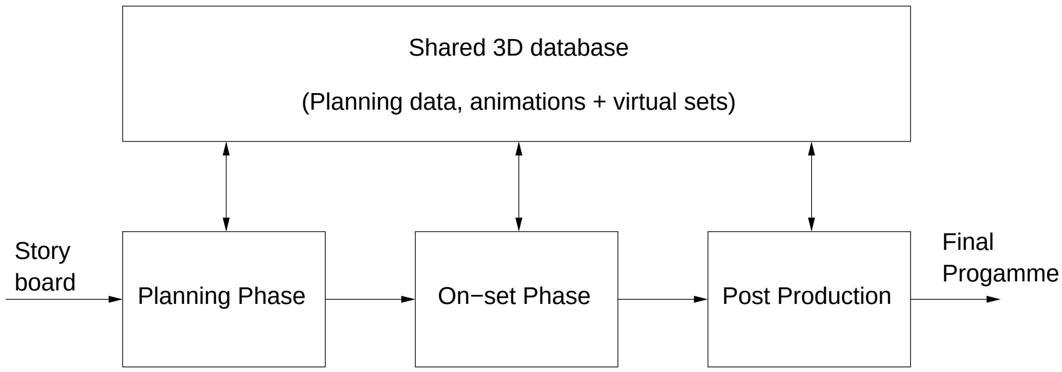

Figure 2. New production work flow involves access to 3D data in each production phase.

3.1. On-set phase in the studio

The 3D data database (including animations) from the planning phase is used and altered in the studio. The background models and the prepared animations are used for pre-visualisation. Two kinds of pre-visualisation modules are implemented in the ORIGAMI studio system:

- In the *director's view* the background scenery, the 3D animation and a preview-quality dynamic 3D model of the actor is rendered onto a monitor as a preview of the final composited programme. This preview enables the director to decide on the exact positions of actors and (virtual or real) props[†]. That means the data set from the planning will be refined and edited during this stage.
- The *actor feedback* is a view-dependent projection onto the walls and floor of the studio. It allows the actor to interact with the virtual world.

The dynamic 3D models of the actor are created with the visual hull reconstruction technique, as discussed in the previous section. The images from the fixed capture cameras are digitised and stored at full frame rate for later off-line processing. The resulting real-time 3D model of the actor is then rendered together with the virtual background of the set and any 3D animations. Any changes in positions of the props or the actors or the camera framing can be made with the 'director's view' tool and are stored into the shared 3D database for later use in post-production.

The immersive actor feedback system works in parallel to the real-time modelling system. It is based on standard data projectors driven from rendering PCs that project onto the special retro-reflective cloth without interference with the chroma-keying facility, since the light from the LED rings dominates the projector light. The head position of one actor is derived from the captured 3D model and used to synthesise view-dependent images of the virtual scene for each projector. In fig. 3 the system is shown during a demo production that was set in the entrance hall of the British Natural History Museum in London. The boy was watching a Pterosaur flying through the entire length of the hall. Four projectors were used to generate this 180 ^{o} view of the hall and the animation.

Each projector is driven by a separate PC. The entire system is distributed and build upon a CORBA communication layer. Details of this architecture can be found in.[8]

3.2. Post-production phase

In the post-production phase the programme is rendered in final quality. That involves final adjustments of the 3D animations using the actual camera footage and the dynamic 3D actor models. Due to the fact that the animators can use the shared 3D database that has already built up through the planning and on-set phase the animation work is much more effective than starting from scratch as is the case in a conventional production flow.

[†]Props are additional items, like furniture on set

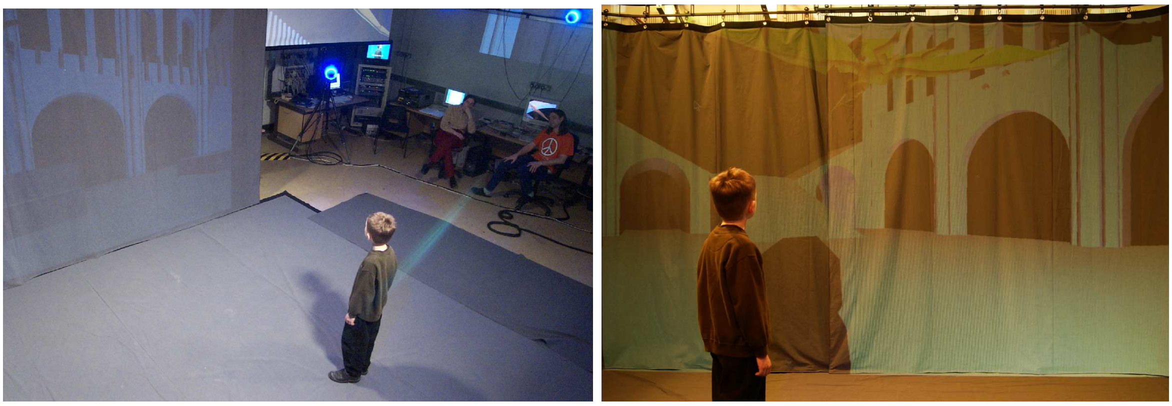

Figure 3. Projection system.

4. RESULTS

The techniques discussed in this paper have been tested in a demo production. For this purpose 3D models of the entrance and the main hall of the Natural History Museum in London were produced by one of the ORIGAMI partners (MIP University of Kiel). These have been used for planning and as background models in a few final scenes of a short demonstration video.

The foreground action was recorded in the experimental studio of BBC R&D using the techniques described in this paper. Fig. 4 shows one frame of the final demonstration video showing a boy watching a Pterosaur, animated by another project partner (Framestore CFC) flying through the entrance hall. Due to the pre-visualisation technique presented in section 3.1 (see fig. 3) the boy was able to keep perfect eye-contact with the dinosaur in all takes. The director of the production could select the take based purely on artistic reasons.

Figure 4. Two frames from the final quality demonstration video.

The figs. 5 and 6 show results of the 3D visual hull reconstruction. In fig. 5 (left) a conventional hierarchical octree-based method was applied to the silhouettes from 12 cameras. The voxel resolution was 128x128x128 or approximately 3 cm voxel length. The generated 3D surface description consists of nearly 17,000 triangles.

The picture in the middle of fig. 5 shows the result of our new super-sampling (one additional super-sampling level) technique that provides the same voxel grid resolution and the roughly the same number of triangles. It can be observed that the reconstruction is smoother and the back-projected model fits better to the original silhouette.

Figure 5. Results of the visual hull computation using standard octree (left), super-sampling (middle) and additional smoothing (right).

The right picture of fig. 5 shows the super-sampled reconstruction applied with a Gaussian smoothing and triangle reduction to 3000 triangles. This model was then textured and imported to a standard animation package (Softimage|XSI) to render the final image quality, including shadows, as depicted in fig. 6.

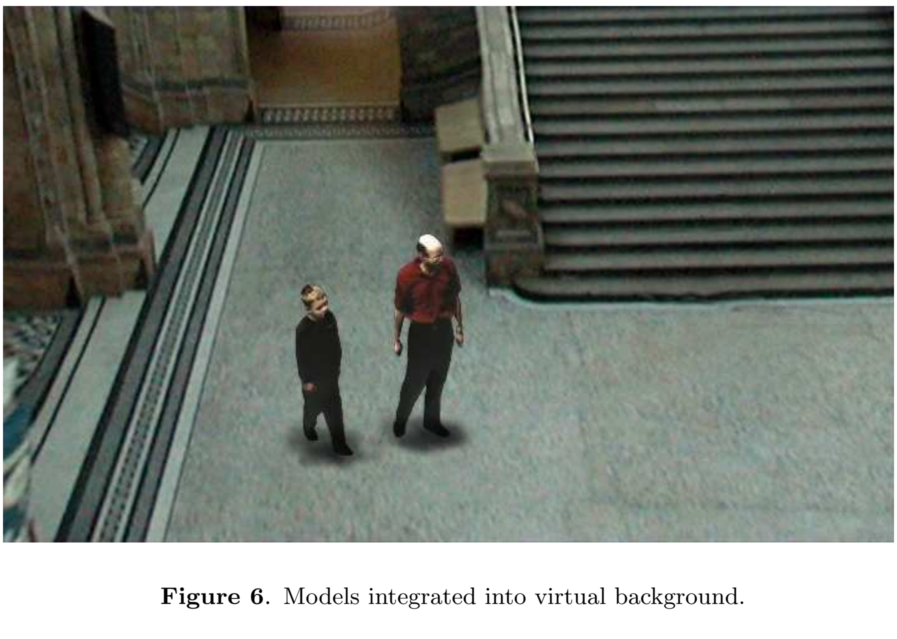

Figure 6. Models integrated into virtual background.

5. CONCLUSIONS

This contribution described techniques for 3D modelling and sensing of dynamic studio scenes and different visualisation techniques for the production of special effects. These were then used in an improved production pipeline.

The main concept of the new production pipeline is the use of a shared 3D database that is used and edited through the different production stages. The use of 3D data in the planning phase is already useful, because it allows the simulation of different camera perspectives and positions of actors and props.

For the on-set phase 3D animations are added. The new feedback system for the director allows decisions to be made on-set and the 3D database to be further refined. The actor feedback module gives the performing artists a fully bi-directional interaction with the virtual world and solves current weaknesses of virtual production techniques, like the important eye-line problem.

A recent demo production has shown the value of these tools for the on-set work and also showed that the post-production work benefits from the acquired 3D data.

Moreover the dynamic 3D modelling techniques developed have been used in post-production to achieve new effects, like full optical integration of real actors with virtual backgrounds (shadow casting and reception) and fully free selectable camera viewpoint (virtual crane moves or 'The Matrix' style effects).

Although the usefulness of a 3D production pipeline has been shown, there is currently no integrated framework that allows the 3D data set to be passed through the entire production chain. Instead a number of separate tools exist. A better integration of the different production phases has to be addressed by future work.

ACKNOWLEDGMENTS

The author would like to thank the contributing colleagues from the BBC R&D department and the ORIGAMI project partners, in particular the the Multimedia Information Processing group of Christian-Albrechts-University of Kiel, Framestore CFC and the Image and Sound Processing Group of Politecnico di Milano for providing the 3D models and animations used for this work.

The work presented in this paper was funded by the European project IST-2000-28436 ORIGAMI.

REFERENCES

1. O. Grau, R. Koch, F. Lavagetto, A. Sarti, S. Tubaro, and J. Woetzel, "The origami project: Advanced tools for creating and mixing real and virtual content in film and tv production," *submitted to IEE Proceedings Vision, Image and Signal Processing* .
2. O. Grau and et al., "New production tools for the planning and the on-set visualisation of virtual and real scenes," in *Conference Proc. of International Broadcasting Convention*, (Amsterdam, NL), September 2003.
3. S. Rosenthal, D. Griffin, and M. Sanders, "Real-time compter graphics for on-set visualization: "a.i." and "the mummy returns"," in *Siggraph 2001, Sketches and Applications*, 2001.
4. Tzidon and et al., "Prompting guide for chroma keying," *United States Patent* , 23 March 1996.
5. C. Cruz-Neira, D. Sandin, T. DeFanti, R. Kenyon, and J. Hart, "The cave: Audio visual experience automatic virtual environment," *Communications of the ACM* **35**, pp. 65–72, June 1992.
6. M. Gross, S. Wuermlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. V. Gool, S. Lang, K. Strehlke, A. V. Moere, and O. Staadt, "blue-c: A spatially immersive display and 3d video portal for telepresence," in *Proc. of ACM SIGGRAPH 2003*, pp. 819–827, (San Diego, USA), July 2003.
7. W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-based visual hulls," in *Siggraph 2000, Computer Graphics Proceedings*, K. Akeley, ed., pp. 369–374, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.
8. O. Grau, T. Pullen, and G. A. Thomas, "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**, pp. 370–380, March 2004.

9. O. Grau, "3d sequence generation from multiple cameras," in *Proc. of IEEE, International workshop on multimedia signal processing 2004*, (Siena, Italy), September 2004.
10. M. Potmesil, "Generating octree models of 3D objects from their silhouettes in a sequence of images," *Computer Vision, Graphics and Image Processing* **40**, pp. 1–29, 1987.
11. R. Szeliski, "Rapid octree construction from image sequences," *CVGIP: Image Understanding* **58**, pp. 23–32, July 1993.
12. W. Niem, "Robust and fast modelling of 3d natural objects from multiple views," in *SPIE Proceedings, Image and Video Processing II*, **2182**, pp. 388–397, (San Jose), February 1994.
13. W. Martin and J. K. Aggarwal, "Volumetric descriptions of objects from multiple views," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5**, pp. 150–158, March 1983.
14. C. R. Dyer, "Volumetric scene reconstruction from multiple views," in *Foundations of Image Understanding*, L. S. Davis, ed., pp. 469–489, Kluwer, Boston, 2001.
15. W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 163–169, ACM Press, 1987.
16. O. Grau and A. Dearden, "A fast and accurate method for 3d surface reconstruction from image silhouettes," in *Proc. of 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 395–404, (London, UK), April 2003.