THE QUEEN'S AWARDS FOR ENTERPRISE 2001

R&D White Paper

WHP 035

September 2002

Use of image-based 3D modelling techniques in broadcast applications

O. Grau *and* G.A. Thomas

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 035

Use of image-based 3D modelling techniques in broadcast applications

Oliver Grau, Graham A. Thomas

Abstract

Conventional video- and audio-based content is still the mainstream media for the broadcast industry. However, there is an increasing demand to use 3D models, mainly in special effects or visualisation of virtual objects, like buildings that no longer exist. On the other hand there are several established or emerging technologies that make explicit use of 3D models and are of interest for the broadcast industry. These applications include: 3D games, 3D internet applications and 3D-TV.

This paper discusses several techniques and systems that allow cost effective creation of this content. The approaches can be classified by the class of shape quality they deliver: 2D-Sprites, incomplete 3D (2.5D) and complete 3D models. This contribution describes an approach for classifying these shape types. An overview of requirements for the generation of object-based 3D models in a number of different applications is also given.

This document was originally published in 2002 Tyrrhenian International Workshop on Digital Communications (IWDC 2002), September 8th - 11th, 2002, Capri, Italy.

Key words: Virtual Production, Virtual Studio, Special Effects, 3D Reconstruction

White Papers are distributed freely on request.

Authorisation of the Head of Research is required for publication.

Use of image-based 3D modelling techniques in broadcast applications

Oliver Grau, Graham Thomas

BBC R&D
Kingswood Warren, Tadworth, Surrey, KT20 6NP, UK

Email: *oliver.grau | graham.thomas@rd.bbc.co.uk*, Web: *http://www.bbc.co.uk/rd*

ABSTRACT

Conventional video- and audio-based content is still the mainstream media for the broadcast industry. However, there is an increasing demand to use 3D models, mainly in special effects or visualisation of virtual objects, like buildings that no longer exist. On the other hand there are several established or emerging technologies that make explicit use of 3D models and are of interest for the broadcast industry. These applications include: 3D games, 3D internet applications and 3D-TV. This paper discusses several techniques and systems that allow cost effective creation of this content. The approaches can be classified by the class of shape quality they deliver: 2D-Sprites, incomplete 3D (2.5D) and complete 3D models. This contribution describes an approach for classifying these shape types. An overview of requirements for the generation of object-based 3D models in a number of different applications is also given.

1 INTRODUCTION

The main use of 3D models in broadcast applications at the moment is in virtual studios and special effects. These methods are often referred to as virtual production. The key idea of virtual production is the composition of virtual and real scene elements or different virtual elements. For this composition several optical phenomena must be harmonised between different kinds of media or components from different sources. The most important phenomena are: (1) match of the camera perspective of 2D footage with rendered virtual components, (2) occlusions, (3) establishing proper lighting and shading of the different scene components and (4) reflections.

In a virtual studio one or more actors are captured in a studio using a camera fitted with a tracking system. The parameters of the real camera are then transfered to a virtual camera that renders the background of the scene. Finally the actor is keyed into this image using the chroma-keying technique. The method is effective but often lacking a sufficient degree of optical interaction. So the example of figure 1 does not show the correct lighting on the (real) actor and a proper shadow on the virtual floor.

In order to create a full and realistic looking optical interaction a 3D description of the real and virtual scene must be available for rendering the synthetic image. This paper investigates several image-based 3D modelling techniques for the use in virtual production. The quality requirements of these 3D models vary depending on the application. Section 3.2 gives a summary of different applications and their requirements.

Figure 1: Virtual studio: Scene of the BBC's programme "Tomorrow's world". Optical interactions are restricted.

The following section gives an overview about related 3D reconstruction techniques.

Section 4 describes approaches and their implementation in the broadcast context. The article finishes with some results and conclusions.

2 RELATED WORK ON SHAPE RECONSTRUCTION

In recent years, a lot of research has been carried out in the field of 3D scene reconstruction and especially in the development of 3D depth sensors. This can be divided into active and passive approaches.

Active 3D sensors use an active illumination technique and are usually more robust compared, for example, to passive stereo. Products using active techniques such as lasers or structured light are available from Cyberware (USA), Wicks & Wilson (UK), Vitronic (Germany) and others. Unfortunately none of these systems are able to capture a moving 3D scene at normal video frame rate. A promising system is the Z-Cam from 3DV Systems (Israel), which uses a special TV camera with video frame rate and a depth sensing method based on time-of-flight of light.

Passive 3D sensors generally use a two-camera rig in conjunction with a stereo vision approach (for example [1, 2, 3, 4]). These systems are not currently in common use in a studio environment, mainly due to their restrictions in accuracy and robustness.

There are also passive methods using more then two cameras. A promising approach is the shape from silhouette method [5, 6, 7]. This has been shown as a fast, scalable and robust method for reconstruction of 3D information from silhouettes, that can be computed easily in real-time in a specially equipped studio using chromakey techniques. In order to use the method the cameras must be calibrated. That means the internal and external camera parameters must be known. These are measured using a calibration pattern and a calibration procedure to estimate the camera parameters [8, 9]. Then the intersection of the bounding volumes of each silhouette is computed. From the remaining convex hull in the form of a voxel representation, a surface description is computed. In the last step a texture map is created using the information from the camera images.

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 [10, 11] makes use of the colour information, i.e. the differences between the generated model and the camera images.

Another advanced approach, that fundamentally makes use of silhouette information is the incorporation of high-level generic models of human bodies [12, 13]. These methods give a good appearance, but are not intended for modelling a person photo-realistically at video frame rate, because they grab a texture map only once. The main application field of these methods is for online applications with a limited bandwidth.

In recent years the techniques used for virtual production have been influenced by image-based methods and plenoptic modelling. The original work on plenoptic modelling used little or no geometry. Meanwhile there is a spectrum of methods known in the literature that are using different qualities of geometrical description. E.g. view-dependant texture-mapping [14], surface-light-fields [15] and the unstructured lumigraph [16]. Although these methods are very promising for photo-realistic modelling, even of very complex environments, they are not well suited for inserting or exchanging objects from or into other environments or changing the lighting of the scene.

3 Applications and Requirements

The requirements for a 3D reconstruction of objects depend both on the application and the kind of optical phenomena that should be realised. The most important *optical phenomena* concerning virtual production are:

a. camera perspective

b. occlusions

c. depth perception

d. shadows

e. light reflections

The *degree of optical interaction* gives a measure of how good the composited scene of virtual and real components realises these phenomena between the inserted objects. The relative importance of these phenomena and the feasibility of implementing them varies with the application. The following sections give a summary of applications in the broadcast context and criteria for the choice of 3D quality for the particular purposes.

3.1 Current and future use of 3D techniques in virtual production

Conventional video- and audio-based content is still the mainstream media for the broadcast industry. However, there is an increasing demand to use 3D models, in virtual studios, in special effects and for the visualisation of virtual objects.

The term *virtual studio* is usually taken to mean the technique of combining live action in a studio with a keyed background image. The latter is updated in real-time by a computer equipped with sufficiently fast graphics hardware to match the movement of the studio camera. Due to the real-time requirement the optical interactions are quite limited. The minimum is to match the camera parameters using camera tracking. The final compositing works in 2D; 3D data is not used except a rough depth estimate for occlusion masking. The virtual camera cannot be moved to a position different from than that of the real camera; control over lighting is also very poor.

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

Pre-visualisation is a tool to help planning the arrangements on the set during an early state of the production, i.e. position of actors, cameras and lighting. The 3D model must allow these positions to be changed interactively. Pre-visualisation becomes more and more important and is used independently from special effects, i.e. also in conventional productions.

On-set visualisation is mainly intended to be a tool during production, i.e. to give feedback to camera operators, actors and directors. Therefore, it does not have the requirement for the best quality, but should be available in real-time.

3D-TV gives the observer the added value of depth perception. Due to lack of the right 3D-display technology this technique is not implemented for home users yet, but in IMAX cinema it already plays a role.

An important application for the future is *interactive programmes* that make explicit use of 3D models and are of interest for the broadcast industry. These applications include: games, edutainment and 3D Internet applications. Here the user gets the full control over the virtual camera.

Before we discuss the question of the right shape quality or representation, it is important to notice that virtual studio systems and also most special effects today are using 2D compositing for the final integration of virtual and real scenes. In particular conventional virtual studio systems do not allow optical interactions other than

Application	RT	shad.	ref.	lc	cc	Class
Virtual Studio	++	(+)	-	-	-	I
On-set Visuali.	+	-	-	-	-	II
Pre-Visuali.	-	+	+	++	++	III
Special Effects	-	+	+	+	(+)	II-IV
3D-TV	-	+	+	+	-	II
Interactive	-	+	(+)	(+)	++	III

Table 1: Shape quality.

matching of the camera perspective and (on some systems) simple occlusions. For special effects some optical interactions like shadows can be realised by additional alpha mattes that ‘dim’ the real image in shadow areas. For features like the change of virtual viewpoints or full control over the lighting a 3D description of the virtual and the real scene is needed.

3.2 What shape quality is needed for which application ?

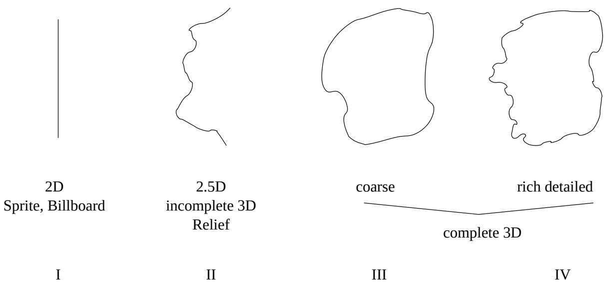

Figure 2: Different classes of shape quality.

Figure 2 gives a classification of different shape qualities used in virtual production [17]. Which one to choose depends strongly on application requirements, like the question of whether the created model is used for real-time rendering or not. Moreover, the shape quality influences the degree of optical interaction that is possible. For example a sprite (class I) when used in a 2D layered system can realise simple occlusion effects, like rendering the sprite object on top or behind another layer, but they do not give fine control, like pushing only a hand through a virtual object.

Table 1 gives a summary of the optical phenomena that are used in the applications mentioned in the previous section and the shape class that should be used. The following abbreviations are used in the header of the table: *RT* = real-time reconstruction and synthesis, *shad.* = shadows, *ref.* = reflections, *lc* = lighting control and *cc* = control over virtual camera. Since the right perspective and occlusions are always needed, these are omitted in the table. A ’-’ in the table indicates that this feature is not important, a ’+’ and ’++’ indicates it is important and very important.

Table 1 shows a clear correlation between the shape quality class and the number of optical phenomena that must be realised. The rich detailed 3D class allows the widest applications, but is not always needed or possible due to the real-time constraints, like in the virtual studio or on-set visualisation.

Special effects cannot be clearly categorised because they differ from case to case. Usually the best class (IV) might be desired, but is not always possible to be realised due to budget restrictions or may not actually be necessary at all.

Interactive applications usually require a complete 3D description, because of the fact that the user has full control over the virtual camera. Due to limitations in resources, namely bandwidth and rendering speed, they will be normally restricted to a coarser and therefore more compact shape description, i.e. class III.

4 APPROACHES

This section gives an overview about 3D reconstruction techniques developed by the BBC. The techniques are developed for use in a studio environment like in figure 3. An exception is the use of the stereo methods, as described in section 4.2, that can be used in a range of different environments.

Figure 3: Studio setup with different capturing systems

4.1 2D- Sprite and Billboard models

The simplest 3D description of an object, according to figure 2 is a 3D polygon, texture-mapped with the camera image. These ‘sprites’ are using alpha-masks to represent arbitrarily-shaped 2D objects and can be created automatically by re-projecting the image frame from the camera position as a polygon into space. The position of this polygon should coincide with the position of the actor in the studio reference co-ordinate system. Therefore, the position and orientation of the camera and the position of the actor must be known.

A camera tracking system, and an actor tracking system using an auxiliary camera, as depicted in figure 4, were developed by us as components in a virtual studio system [9, 18]. The optical interaction is limited here to simple occlusions. The virtual camera can be moved away from the real viewpoint only if the viewing angle is not significantly changed (see figure 7 + 8).

An extension of the planar sprite model was proposed by Grau, et. al in [17]. This method is based on the idea that objects are usually more elevated in the middle. This assumption can be used, as depicted in figure 5, by computing the 2D distance function (figure 5 c) of the 2D silhouette (figure 5 b) and a polygonal elevation function (figure 5 d).

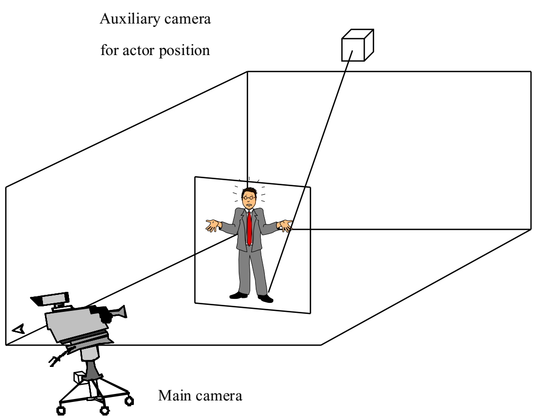

Figure 4: The position of the actor is tracked by an auxiliary camera

Figure 5: Shape from one silhouette. a) camera image, b) object mask, c) 2D distance function, d) 3D mesh.

4.2 Incomplete 3D (Relief)

Within the EU-funded project MetaVision [19], we are studying applications that can be supported by the "2.5D" shape class of figure 2, and developing a method to capture such data. The main application we are considering is the composition of 3D objects into an image sequence using a depth map of the sequence to support optical interactions such as occlusions and shadows between real and virtual elements. Other potential applications are simulations of effects such as depth-of-focus and fog, and the generation of stereoscopic image sequences.

We have chosen to develop a multi-baseline stereo system, with an auxiliary camera positioned either side of the main camera. The main camera could be any standard or high-definition TV or film camera, whereas the auxiliary cameras are lower-cost monochrome cameras with conventional TV resolution. It is not practical to use high resolution colour cameras for the auxiliary cameras due to constraints of size, cost and data capture bandwidth. In our current experimental set-up, all three cameras are 576-line cameras and are running synchronised at 25Hz with progressive scanning.

We adopted a passive multi-baseline stereo approach, instead of an active method, so that the system would be usable outdoors and over relatively long ranges. By adjusting the baseline between the cameras, the practical usable range can be 20m or more, although depth resolution decreases significantly with distance. The use of two auxiliary cameras allows us to obtain a depth map for the central camera with significantly fewer occlusions than with a single auxiliary camera. Occlusions inevitably lead to uncertainties in the computed depth around object boundaries, which can be a particular problem for applications such as segmentation, where a reliable depth estimate at object boundaries is important.

As we are mainly considering applications such as special effects implemented in post-production, there is no requirement for a real-time depth image, so we can use non-real-time depth estimation software. However, we cannot afford to use a highly computationally-intensive algorithm, since we need to process sequences containing many thousands of images in a length of time acceptable to typical users of post-production software. A processing time of a few seconds per image on a modern PC is likely to be acceptable.

We have developed a block-based multi-baseline disparity estimator, that seeks to minimise the match error between each block in the central camera image and corresponding regions of the images from the auxiliary cameras. We use spatial and temporal recursion to form an initial prediction of the disparity, both to increase processing speed and ensure a smooth disparity field. To provide the pixel-level disparity field that most applications require, we refine the block-based field by testing disparities for each pixel using candidates from the block containing the pixel and the four immediately adjacent blocks, and measuring the match error over a small region centred on the pixel. Further details may be found in [20].

4.3 Complete 3D

Within the EU-funded project ORIGAMI [21, 22] the BBC is developing a studio-based system that allows the generation of complete 3D models of dynamic scenes both in real-time and offline. ORIGAMI addresses full 3D scene composition for achieving full optical interaction between real and virtual scenes. Therefore the real scene elements are virtualized. The 3D description is then imported into a commercial 3D animation package together with the virtual scene components. The optical interaction is established just by using the render functionality of the animation package that copes inherently with occlusions, shadows and reflections.

Current experiments are using 6 fixed, calibrated cameras and one moving camera equipped with the *free-d* camera tracking system developed by the BBC. The studio is equipped with a special retro-reflective cloth. Around the lens of each camera is a ring of blue LEDs. The light reflected from the retro-reflective cloth allows a robust chroma key of the images.

The real-time shape reconstruction is used for on-set visualisation, which is integrated into the studio system. It gives feedback to actors, camera operators and the director. The actor feedback is realized by generating view-dependent images and projecting them onto the studio floor and walls. On the other hand there should not be any light projected on the actor. Therefore the area where the actor stands is masked out using a 3D model of him created in real-time.

The feedback for camera operators and directors is given on a computer screen. It uses a texture-mapped

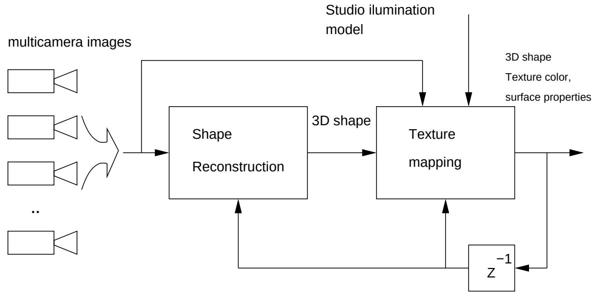

Figure 6: Generation of dynamic 3D models of actors in ORIGAMI

version of the 3D model used for the projection mask.

For the basic shape reconstruction of the scenes in the studio the shape-from-silhouette approach [5, 6, 7] is used, because it makes use of the available chroma key facility.

Currently, initial experiments with shape-from-silhouette for the shape reconstruction are being carried out. First results are presented in section 5. The shape of the objects is represented as 3D triangular meshes. To create a coloured model a texture map is computed, as depicted in the functional block diagram in figure 6. In order to be able to use the generated 3D models in a different illumination situation the texture maps contain diffuse and reflective surface properties. For the computation of these parameters the lighting situation of the studio is measured and used. An algorithm that makes use of this information to "un-light" the texture maps is under development.

An important requirement is that the techniques developed can be applied to dynamic objects and the final result is a sequence of images. Therefore it is important that the generated 3D description is also consistent over time, that means the final sequence should not show significant temporal artefacts. In order to cope with this requirement a recursive approach will be developed that makes use of the previously computed 3D shape and texture. This is indicated by the signal feedback in figure 6.

5 Results

This section presents some results of the approaches described in section 4.

Figure 7 shows an experiment with textured polygons. The original idea was to create a simple 3D description that can be used as a rough actor model in a MPEG-4 system[23].

The system also allows the movement of the virtual camera to a viewpoint different from that of the real camera. This gives reasonable results only if the virtual camera keeps roughly the same viewing angle to the scene as the real camera. If the virtual camera is moved around the scene object, as depicted in figure 8, the limitations of this shape representation become obvious.

An extension of the textured polygon is an elevation grid using the 'shape-from-one-silhouette' method as described in section 4.1. The resulting shape approximation allows much better optical interactions and in particular the virtual camera can be change to some extent.

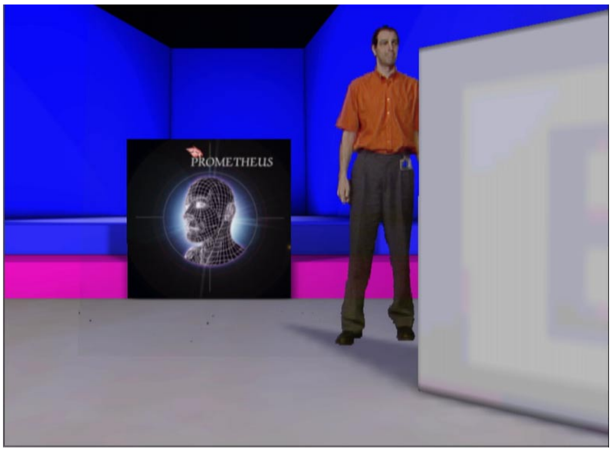

Figure 7: Actor modelled using a textured polygon with alpha-channel

Figure 8: Image rendered form a viewpoint significantly different to the viewpoint of the real camera

Figure 9 show a still image from a sequence[1] created with an animation package. There is a freedom in moving the virtual camera viewpoint of about $\pm 15^\circ$ around original camera viewing angle. Further, the dancer is casting shadows and reflections to the street and is also receiving a shadow.

Figure 10 shows some results from the use of depth information for keying, as described in Section 4.2. The upper image shows one frame from an outdoor sequence. The tree is about 19 m away from the camera. The middle image shows the disparity map for this image, derived from the main image plus two monochrome images from auxiliary cameras positioned 0.25m to either side. The computation time for this disparity image was approximately 7 seconds on a 1.4Ghz Pentium 3 PC, and the disparity range is approximately 5-65 pixels. By thresholding the disparity map at an appropriate value, an alpha signal was created and used to insert a virtual object between the person and the tree, as shown in the lower image. The depth map shows some artefacts around the edges of the person where there is low contrast between the foreground and background, but nevertheless a reasonable key can be achieved. Further work is required

[1] Available on www.ist-metavision.com

Figure 9: Model of a dancer integrated into virtual street model

to improve the disparity map, particularly in low-contrast areas.

Figure 11 shows a visualisation of a multi-camera system as developed within the ORIGAMI project. It shows the original 6 cameras and a number of 3D models from a captured sequence created using the shape-from-silhouette algorithm as described in section 4.3. An example of the use of these models in a virtual environment is depicted in figure 12.

6 Future Work and Conclusions

This contribution discussed the use of 3D modelling techniques for broadcast applications. Recent applications are mainly focused on the integration of virtual and real scenes for TV production, as in virtual studios and special effects. Further 3D techniques are used to support the production process itself by pre- and on-set visualisation. In the future the use of 3D media, as in internet services and games might become an additional application area.

Section 3 gave an analysis of the requirements of the application areas on one side and proposed four different classes of shape quality on the other end. Useful criteria for the choice of the shape class are the degree of optical interaction that is needed, the degree of freedom of the virtual camera and the virtual lighting.

The approaches discussed in section 4 give a selection of methods implemented mainly for the use in a studio environment. The results give some examples of the quality of the optical interactions under the given constraints.

The technique using just a texture-mapped polygon is quite restricted in usability. The extension of that approach, that creates an elevation grid by using the 2D silhouette of an actor is surprisingly efficient. It provides limited support for re-lighting the scene and creating shadow effects. Furthermore it allows the virtual viewpoint to be moved over a limited range.

In order to give more freedom in the variation of the virtual camera and lighting, a multi-camera approach is investigated, which is currently under development in the IST ORIGAMI project. First results are quite promising. Next steps in the development are dedicated to the reconstruction of proper texture maps, that take colour and reflection features of the surface.

The use of a depth map provides support for a range of applications where the viewpoint of the camera remains essentially fixed, but where effects such as occlusions of inserted virtual objects by real scene elements need to be handled. Initial results show promise, but have highlighted problems with real-world scenes, particularly where areas of low contrast make image segmentation difficult. Further work is planned in this area.

References

[1] M.J.P.M. Lemmens, "A survey on stereo matching techniques," in *International Archives of Photogrammetry and Remote Sensing*, 1988, vol. 27, pp. 11–23.

[2] L. Falkenhagen, "Depth estimation from stereoscopic image pairs assuming piecewise continuous surfaces," in *Image Processing for Broadcast and Video Production*, November 1994, pp. 115–127.

[3] Manfred Ziegler and Lutz Falkenhagen et al, "Evolution of stereoscopic and three-dimensional video," *Signal Processing: Image Communication*, vol. 14, pp. 173–194, 1998.

[4] J.-R. Ohm et al., "A realtime hardware system for stereoscopic videoconferencing with viewpoint adaptation," *Signal Processing: Image Communication*, vol. 14, pp. 173–363, 1998.

[5] M. Potmesil, "Generating octree models of 3D objects from their silhouettes in a sequence of images," *Computer Vision, Graphics and Image Processing*, vol. 40, pp. 1–29, 1987.

[6] Richard Szeliski, "Rapid octree construction from image sequences," *CVGIP: Image Understanding*, vol. 58, no. 1, pp. 23–32, July 1993.

[7] Wolfgang Niem, "Robust and fast modelling of 3d natural objects from multiple views," in *SPIE Proceedings, Image and Video Processing II*, San Jose, February 1994, vol. 2182, pp. 388–397.

[8] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE J. Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[9] Graham A. Thomas, J. Jin, T. Niblett, and Urquhatr, "A versatile camera position measurement system for virtual reality tv production," in *Conference Proc. of International Broadcasting Convention*, Sept. 1997.

[10] S. Seitz and C. Dyer, "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 151–173, 1999.

[11] K. Kutulakos and S. Seitz, "A theory of shape by shape carving," *Intl. Journal of Computer Vision*, vol. 38, no. 3, pp. 197–216, 2000.

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

[13] S. Weik, J. Wingbermühle, and W. Niem, "Automatic creation of flexible antropomorphic models for 3d videoconferencing," *Journal of Visualization and Computer Animation*, vol. 11, pp. 145–154, 2000.

[14] Paul E. Debevec, George Borshukov, and Yizhou Yu, "Efficient view-dependent image-based rendering with projective texture-mapping," in *Proc. of 9th Eurographics Rendering Workshop*, Vienna, Austria, June 1998.

[15] Daniel N. Wood, Daniel I. Azuma, Ken Aldinger, Brian Curless, Tom Duchamp, David H. Salesin, and Werner Stuetzle, "Surface light fields for 3D photography," in *Siggraph 2000, Computer Graphics Proceedings*, Kurt Akeley, Ed. 2000, pp. 287–296, ACM Press / ACM SIGGRAPH / Addison Wesley Longman.

[16] Chris Buehler, Michael Bosse, Leonard McMillan, Steven J. Gortler, and Michael F. Cohen, "Unstructured lumigraph rendering," in *SIGGRAPH 2001, Computer Graphics Proceedings*, Eugene Fiume, Ed. 2001, pp. 425–432, ACM Press / ACM SIGGRAPH.

[17] Oliver Grau, Marc Price, and Graham A. Thomas, "Use of 3d techniques for virtual production," in *Proc. Of SPIE, Conference Proc. of Videometrics and Optical Methods for 3D Shape Measurement*, January 2001, vol. 4309.

[18] Marc Price and Graham A. Thomas, "3d virtual production and delivery using mpeg-4," in *Conference Proc. of International Broadcasting Convention*, Sept. 2000.

[19] "Metavision project website," http://www.ist-metavision.com.

[20] Graham A. Thomas and Oliver Grau, "3d image sequence acquisition for tv & film production," in *Conference Proc. of 1st Symp. on 3D Data Processing Visualization Transmission*, Pavova, Italy, June. 2002.

[21] "Origami project website," http://www-dsp.elet.polimi.it/origami/.

[22] Geovanni Bazzoni, Enrico Bianchi, and Oliver Grau et al., "The origami project: advanced tools and techniques for high-end mixing and interaction between real and virtual content," in *Conference Proc. of 1st Symp. on 3D Data Processing Visualization Transmission*, June. 2002.

[23] M. Price, J. Chandaria, and O. Grau et. al., "Real-time production and delivery of 3d media," in *Conference Proc. of International Broadcasting Convention*, Sept. 2002.

Figure 10: Depth-keying

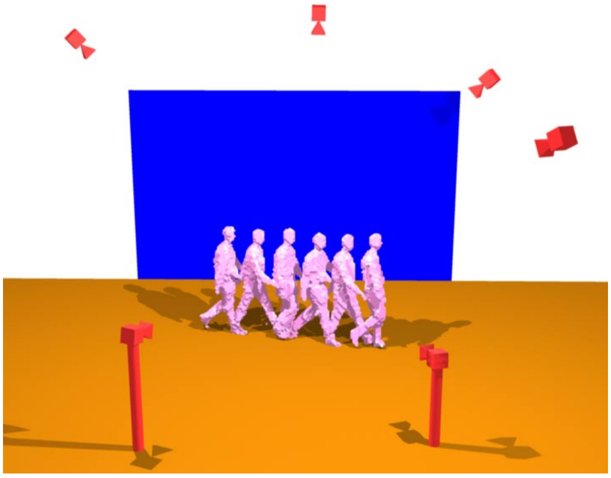

Figure 11: Multi-camera system

Figure 12: 3D model from multi-camera system integrated into virtual scene