R&D White Paper

WHP 033

July 2002

Use of 3-D Techniques for Virtual Production

O. Grau, M. Price *and* **G.A. Thomas**

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 033

Use of 3-D Techniques for Virtual Production

O. Grau, M. Price and G.A. Thomas

Abstract

Virtual production for broadcast is currently mainly used in the form of virtual studios, where the resulting media is a sequence of 2-D images. With the steady increase of 3-D computing power in home PC's and the technical progress in 3-D display technology, the content industry is looking for new kinds of programme material, which makes use of 3-D technology. The applications range from analysis of sport scenes, 3DTV, up to the creation of fully immersive content.

In a virtual studio a camera films one or more actors in a controlled environment. The pictures of the actors can be segmented very accurately in real time using chroma keying techniques. The isolated silhouette can be integrated into a new synthetic virtual environment using a studio mixer. The resulting shape description of the actors is 2-D so far. For the realisation of more sophisticated optical interactions of the actors with the virtual environment, such as occlusions and shadows, an objectbased 3-D description of scenes is needed. However, the requirements of shape accuracy, and the kind of representation, differ in accordance with the application.

This contribution gives an overview of requirements and approaches for the generation of an object-based 3-D description in various applications studied by the BBC R&D department. An enhanced Virtual Studio for 3-D programmes is proposed that covers a range of applications for virtual production.

This document was originally an invited paper presented at SPIE conference on 'Videometrics and Optical Methods for 3D Shape Measurement', 22-23 Jan.2001, San Jose, USA.

Keywords: Virtual Studio, 3-D Reconstruction, Virtual Production, 3-D Programmes, Shape Representation, Shape from Silhouette

White Papers are distributed freely on request.

Authorisation of the Head of Research is required for publication.

Invited paper presented at SPIE conference on 'Videometrics and Optical Methods for 3D Shape Measurement', 22-23 Jan.2001, San Jose, USA.

Use of 3-D Techniques for Virtual Production

Oliver Grau, Marc Price, Graham A. Thomas

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

ABSTRACT

Virtual production for broadcast is currently mainly used in the form of virtual studios, where the resulting media is a sequence of 2-D images. With the steady increase of 3-D computing power in home PC's and the technical progress in 3-D display technology, the content industry is looking for new kinds of programme material, which makes use of 3-D technology. The applications range from analysis of sport scenes, 3DTV, up to the creation of fully immersive content.

In a virtual studio a camera films one or more actors in a controlled environment. The pictures of the actors can be segmented very accurately in real time using chroma keying techniques. The isolated silhouette can be integrated into a new synthetic virtual environment using a studio mixer. The resulting shape description of the actors is 2-D so far. For the realisation of more sophisticated optical interactions of the actors with the virtual environment, such as occlusions and shadows, an object-based 3-D description of scenes is needed. However, the requirements of shape accuracy, and the kind of representation, differ in accordance with the application.

This contribution gives an overview of requirements and approaches for the generation of an object-based 3-D description in various applications studied by the BBC R&D department. An enhanced Virtual Studio for 3-D programmes is proposed that covers a range of applications for virtual production.

Keywords: Virtual Studio, 3-D Reconstruction, Virtual Production, 3-D Programmes, Shape Representation, Shape from Silhouette

1. INTRODUCTION

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 harmonized between different kinds of media or components from different sources. The most important phenomena are: match of camera parameters, in the case of composition of 2-D footage, with rendered virtual components and establishing proper lighting and shading of the different scene components.

Virtual production for broadcast is currently mainly used in the form of virtual studios, where the resulting media is a sequence of 2-D images. With the steady increase of 3-D computing power in home PC's and the technical progress in 3-D display technology, the content industry is looking for new kinds of programme material, which makes use of 3-D technology. The applications range from analysis of sport scenes, 3DTV, up to the creation of fully immersive content.

In a virtual studio a camera films one or more actors in a controlled environment. The pictures of the actors can be segmented very accurately in real time using chroma keying techniques. The isolated silhouette can be integrated into a new synthetic virtual environment using a studio mixer. The resulting shape description of the actors is 2-D so far. For the realisation of more sophisticated optical interactions of the actors with the virtual environment, such as occlusions and shadows, an object-based 3-D description of scenes is needed. However, the requirements of shape accuracy, and the kind of representation, differ in accordance with the application.

As a common framework the typical application scenarios considered in this contribution are typical studio scenes, containing one or more actors, as found in many different types of programmes. The further use of the acquired data and the final representation (e.g. 2-D (Sprite), 2.5-D or 3-D) differs and depends on the application.

Further author information: E-mail: (Oliver.Grau | Marc.Price | Graham.Thomas)@rd.bbc.co.uk, http://www.bbc.co.uk/rd/

1.1. Related Work on Shape Reconstruction

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

Active 3-D sensors use an active illumination technique and are more robust compared to passive stereo for example. 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 3-D 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 kind of time-of-flight sensing method.

Passive 3-D sensors generally use a two camera rig in conjunction with a stereo vision approach (for example[16,2,4,3]). 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.[7–9] This has been shown as a fast, scaleable and robust method for reconstruction of 3-D information from silhouettes, that can be computed easily in real-time in a number of specially equipped studios using chroma-key techniques. In order to use the method the camera sensors must be calibrated. That means the internal and external camera parameters must be known. That is done by using a calibration pattern and a calibration procedure to estimate the camera parameters.[12,1] Then the intersection of the bounding volumes of each silhouette is computed. From the remaining convex hull in 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 approach[14,15] 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.[17,18] These methods give good appearances, but are not intended for modelling a person photorealistically 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.

The former approaches are object based. A new class of representations and approaches known as image-based methods have been proposed recently. The most prominent are: Plenoptic modelling[19] and light-fields or lumigraphs.[20,21] These methods are an extension of simple image-based methods, such as texture maps. They do not make use of an explicit shape representation at all and they can cope very well with reflections and other lighting effects. Although these methods are very promising for photorealistic modelling, even of very complex environments, they are not well suited for inserting or exchanging objects from or into other environments. Because this is the basic idea of virtual production, in the following sections only object-based representation for the objects are considered.

This following section gives an overview of requirements in various applications that we have studied. Section 3 describes some approaches for the generation of object-based 3-D description in detail. The paper concludes with some results and conclusions.

2. REQUIREMENTS AND APPLICATIONS

The requirements for a 3-D description or representation of object shapes depend on the given application. This section first gives a brief overview of shape representations classes. Then different applications and their specific requirements are described in more detail. Most of the applications are in the field of entertainment and in the context of this paper all applications considered include temporal changes of the components, including changing camera parameters.

The requirements for shape representation depend mainly on the kind of optical phenomena that should be realised within the different applications. The most important optical phenomena concerning virtual production are:

a. perspective projection described by parameters of a moving camera

b. occlusions of scene components by other components

c. depth perception

d. shadows

e. light reflections

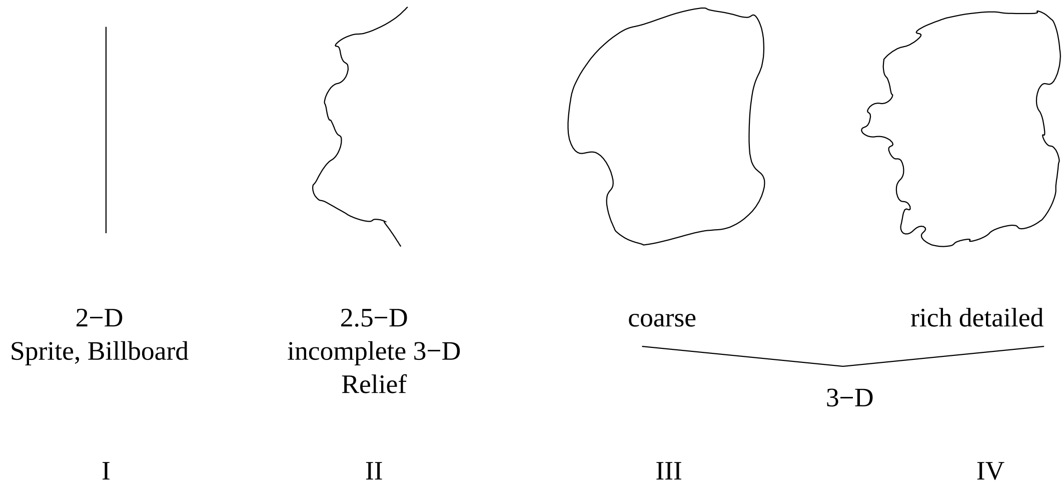

Figure 1. Different classes of quality in shape representation

The relative importance of these phenomena and the feasibility of implementing them varies with the application.

Several classes of object-based shape representation are in use for virtual production as depicted in Fig. 1. The simplest class is just a planar polygon or sprite in space; that means this representation has no depth extension. The other classes (II-IV) can cope with depth information. In addition to the shape, the surface colour is usually captured and stored in a texture map that is assigned to the geometrical shape description.

In the following sections typical applications using virtual production are discussed.

2.1. Virtual Studio

The term virtual studio is usually taken to mean the technique of combining live action in a studio with a keyed background image, where the background image is updated in real-time to match the movement of the studio camera. This requires two additional devices in addition to the equipment normally required for chroma-keying: a device for measuring the motion and lens angle of the camera, and a means of generating a background image corresponding to the camera's viewpoint. A video delay unit (typically providing a few frames of delay) is also needed to compensate for the delay in generating the background image. This is used to delay the camera's video signal before it is keyed with the background image.

Most virtual production systems can be classified as either “2-D” or “3-D”.

In a *2-D system*, the camera movement is limited to pan, tilt and zoom, in exchange for significant simplification of the system. Measuring such constrained camera movement is much simpler than measuring the movement of a ‘free’ camera; for example, rotary encoders can be attached to the camera mounting to measure the rotation relative to a fixed pedestal. Ideally, the camera should be mounted so that the nodal point of the lens remains fixed. Normal camera mounts tend to place the pivot point some distance behind the lens, and a special mounting plate may be necessary to allow the camera to be mounted further back.

Since the position of the camera remains fixed, there are no parallax effects in the background, so the background may be treated as a 2-D image (usually flat, but panoramic images may also be used in some systems). A device similar to a conventional DVE (Digital Video Effects device) may be used to apply the required affine transform to the background image in accordance with the measured pan, tilt and zoom values from the camera. The quality of the background image from such systems can be truly photorealistic, limited only by the resolution of the original image and the quality of the interpolation used in the affine transform that maps the image to match the camera view. Background images may be derived in many ways, including still photographs of real scenes, rendered views of 3-D models, or video signals, although care needs to be taken to match the lens angle and camera height to that of the camera in the studio. The limitation of camera movement reduces the flexibility of the technique somewhat, although in practice many programmes do not require camera translation whilst the camera is ‘on air’.

A *3-D virtual production system* allows translational camera movement in all three dimensions, as well as pan, tilt, roll and zoom. This necessitates the use of a more complex camera tracking system, as well as a real-time 3-D renderer to generate the background image.

Several kinds of camera tracking system are available. Some are mechanical, relying on robotic camera pedestels or cameras mounted on rails. Others rely on image analysis to identify markers placed either in the scene or on the ceiling,[1] or use arrangements of markers on the camera that are viewed by several cameras looking down into the studio. A highly-accurate tracking system is needed to ensure negligible 'drift' between virtual and real parts of the scene; typical requirements are to measure position to 1 mm, and direction-of-view to 0.01 °.

The need to render a 3-D scene in broadcast-quality, at a constant rate of 50 or 60 images per second, generally requires the use of a very high-end graphics computer, such as an SGI Onyx, although systems based on PCs are starting to become usable. A carefully-designed model must also be used, which does not exceed the capacity of the rendering system in terms of polygons, texture maps, fill rate, and so on.[10] An alternative to rendering the broadcast-quality image in real-time is to render a real-time lower-quality image (sufficient for a cameraman to use for framing), and record the camera parameters to control a subsequent off-line 'post-rendering' operation. This allows cheaper computers to be used and increases the quality of the end result, in exchange for a significant amount of post-production time.

Some systems are available which combine elements of both the 2-D and 3-D approaches. For example, by passing the image from the studio camera through a DVE, it is possible to simulate some kinds of camera movement, allowing the scene to be viewed from a 'virtual' camera position that differs from that of the real camera in the studio. However, the virtual viewpoint should not be too far away from the real viewpoint, or the actor(s) will appear distorted.

It is worth noting that a 3-D virtual studio does not generate a 3-D result: the final image is a conventional 2-D TV picture, and all information on the 3-D nature of the background has been lost in the rendering process. There is no way in which, for example, the 3-D nature of the scene can be exploited in post-production.

2.2. Linear Stereo Programme

Like programmes created with virtual studio techniques, linear stereo programmes do not offer interactivity to the viewer. However, they do add a new quality to the medium: depth perception. Therefore, the physical viewer interface has to be via a stereo display technique. That means that unlike the previous virtual studio application, the output of this one is 3-D. But at the current state of the art the usual approach is by projecting two video channels (one for each eye), as in IMAX 3-D for example, or the stereoscopic virtual studio demonstrated by the ACTS-MIRAGE project.[11] These systems do not provide parallax due to head motion. For a real 3-D implementation, using emerging auto-stereoscopic displays, an explicit 3-D scene description is needed.

Like the virtual studio application the correct camera projection and the occlusion phenomena are important. Also, the viewer is not able to move his viewpoint. Due to the implementation of depth perception a sprite or billboard is only sufficient for objects in the background. For objects in the close range at least a 2.5-D description must be used (object class II, Fig. 1).

2.3. 3-D Interactive Programme

A 3-D interactive programme gives the viewer the ability to move his viewpoint, more or less arbitrarily. This means that a complete 3-D description is needed for objects in the scene. The most prominent applications are: 3-D games and (emerging) 3-D web content.

Currently computer or TV monitors are most commonly used as the display device for these applications. Hence, this means that there is no implementation of depth perception. Highly accurate shadow casting is also not implemented. For these reasons the most common shape representation is a coarse 3-D description (object class III, Fig. 1). This representation gives a good compromise between good visual appearance and the amount of data required for the scene description. It is necessary to limit the quantity of data due to bandwidth limitations in web applications or in the graphics render engine, that creates synthetic images of the 3-D scene in real time.

2.4. Telepresence

Telepresence expands conventional teleconference systems into the 3rd dimension.[4,5] The requirements are similar to linear stereo programmes, due to the fact that the user viewpoint is more or less fixed during a session. A sufficient representation is 2.5-D or a coarse 3-D description (object class II or III, Fig. 1).

An important requirement of this application is that the 3-D shape reconstruction and the creation of the output are in real time. Due to this limitation, most proposed approaches use dense depth maps as the 3-D data representation.

2.5. 3-D Animations (for special effects)

As a part of nearly all recent feature film productions, special effects make manifold use of 3-D data. The most common practice is to add some virtual objects to a scene. In the final result, a 2-D image sequence, these additional objects are integrated to create the illusion that they were in the original scene. To achieve a perfect illusion, not only must camera perspective and occlusion be realised, but also shadow casting between virtual and real scene objects, including any light reflections must be considered. In order to create realistic shadows, 3-D models of both the virtual and the real objects must be obtained.

The appropriate object representation depends on what optical phenomena should be implemented. If for example, a person is captured for insertion into a virtual scene, it is probably sufficient to have a rough 3-D description only (object class III, Fig. 1), for creating shadows on the (virtual) floor. If the optical interactions are more complex, for example, when several different objects are casting shadows or reflecting light onto each other, a rich detailed 3-D description is needed (object class IV). Due to the demanding optical interactions the rendering processes are usually offline, that means non-real time.

2.6. Enhanced Virtual Studio for 3-D programmes

As explained in the above description of Virtual Studios, shape representations of actors and props in the studio are currently only 2-D (class I), and the combination with the background scene (either 2-D or 3-D) takes place in a chroma-keyer. The only situation where some form of 3-D information is sometimes used is in measuring the approximate positions of actors, to control occlusions between actors and virtual elements. Actor positions can be obtained in several ways, for example by using a camera looking down onto the chroma-key studio floor to identify the position of the silhouette of the actor. This can be used to automatically switch mask signals to the keyer.

If it were possible to get more detailed 3-D information relating to the actors and props, then it would be possible to add additional features into a virtual studio system. Some examples are as follows:

- Moving the viewpoint away from that of the real camera: as mentioned previously, a limited degree of movement of the viewpoint away from that of the real camera is possible with 2-D shape information. 'Incomplete 3-D' shape (Class II) would permit a wider range of virtual camera viewpoints (enough to produce a stereoscopic image, for example), but not enough to allow full freedom of viewpoint movement. Full 3-D shape would support arbitrary viewpoints, including views from above and behind.
- Delivery of content in full 3-D form: if 3-D descriptions of actors and props are available, then it becomes possible to deliver the programme as a collection of animated 3-D objects, so that the viewer may choose the viewpoint, and interact with objects in the scene. Additionally, the scene may also be rendered for a 3-D display. Such a programme now resembles a 3-D Interactive Programme as described above, but it has been produced using Virtual Studio techniques, rather than being authored off-line in the manner of a video game. The programme could be delivered using MPEG-4. This possibility is currently being investigated in the PROMETHEUS project.[6]
- Addition of special effects: material captured in a virtual studio that included the creation of 3-D models of actors and props can be enhanced in the same ways as described above for special effects. This could include the addition of virtual shadows and virtual lighting, leading to the more realistic combination of real and virtual elements.

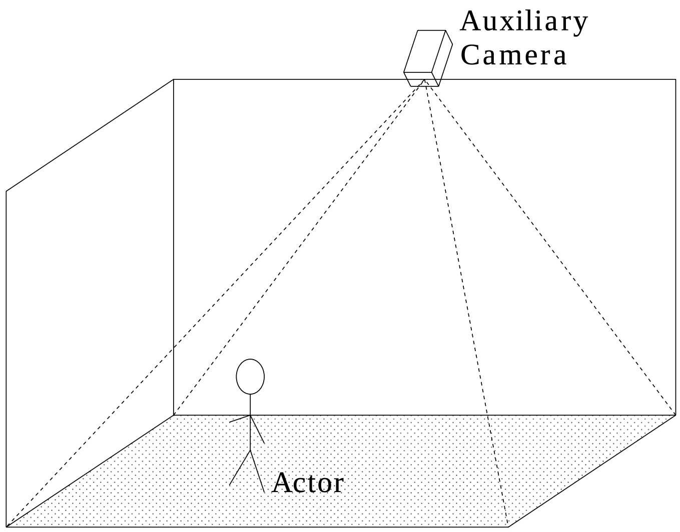

Figure 2. Auxiliary camera for actor tracking

3. SHAPE GENERATION IN AN ENHANCED VIRTUAL PRODUCTION SYSTEM

This section describes approaches for shape capturing, as needed for an enhanced virtual studio system as described in section 2.6.

3.1. The 2-D Sprite Actor Model

A simplified approach to the highly complex problem of realistic actor modelling, is to capture video of the actor with a normal camera, and to texture-map it onto a transparent 2-D object (class I), within the virtual set. The actor-plane is then positioned so as to show the actor standing at the appropriate location. Hence, in addition to the camera video signal, we also need an alpha signal (to key out the real background captured by the video camera), and a real-time actor position signal (to allow the actor model to be appropriately positioned at any point in time).

It is reasonable to assume that the actor performs in a normal virtual studio set-up, with a blue cyclorama background. Hence, the alpha signal can be derived using a conventional chroma-keyer. For our experiments, we have used 'TRUEMATTE', a system based on a retro-reflective cloth with a ring of blue LEDs fitted around the camera lens. This allows greater flexibility with studio lighting, while simultaneously providing the required blue background for correct chroma-keying.

The use of a conventional virtual-studio set-up also enabled us to implement a simple, yet effective means of tracking the position of the actor. The technique uses a small auxiliary video camera, located at a convenient position at the edge of the studio ceiling. The auxiliary camera views down onto the studio floor, as shown in figure 2. With the 'TRUEMATTE' system, a ring of LED's is fitted around the lens of the auxiliary camera, so that the image of the actor appears as a dark silhouette against a bright background. For conventional 'blue-screen' backgrounds, the silhouette image can be formed using a chroma-keyer. The position of the actor on the studio floor is then resolved from the silhouette image by computing the lowest point of the silhouette, (usually corresponding to the feet), and mapping its location to real co-ordinates on the studio floor. The mapping is determined from prior knowledge of position, orientation, and focal length of the auxiliary camera. For larger studios, additional auxiliary cameras would be required.

The size, location, and orientation of the actor model depend on both the location of the actor, and the position, orientation and lens angle of the studio camera viewing the actor. In our experiments, we used the Radamec 'free-d' camera measurement system to provide the required camera parameters. This system uses the methods described in.[1] The optimum orientation of the actor model also depends on other factors, such as the range of 'virtual camera' viewpoints likely to be used to view the virtual scene.

Greatest realism is achieved if the viewpoint is constrained to be equivalent to that of the studio camera. Under these conditions, the realism is maintained by keeping the rotation angle of the actor model perpendicular to the optic axis of the

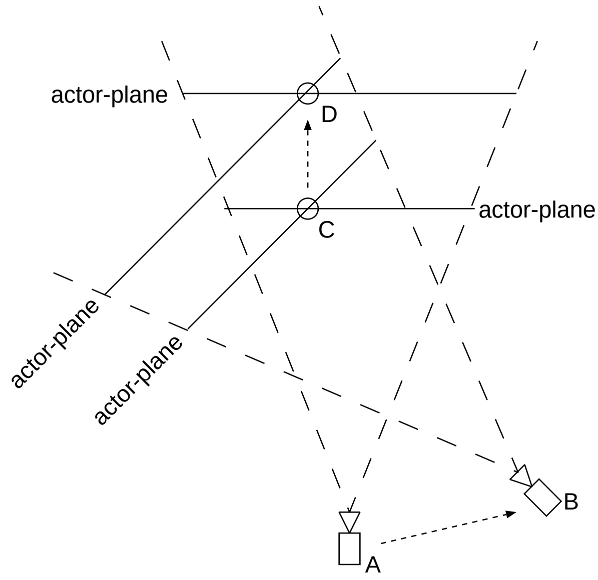

Figure 3. Locating, orientating, and sizing the actor-plane (top view)

studio camera. Using this approach, the actor model is located within the set according to the distance of the actor from the studio camera. Specifically, this is done so that the point where the actor's feet are mapped onto the actor model, corresponds with the actor location, as given by the actor-tracking system. The rotational axis of the actor model is vertical, and intersects the actor's feet. The size of the actor model is adjusted to fill the entire field of view of the studio camera.

This process, which we shall term 'billboarding with respect to studio camera', is illustrated in figure 3, which shows the studio camera moving from point A to point B, and the actor moving from point C to point D.

If, on the other hand, the viewpoint is independent of the studio camera, then the flatness of the actor model becomes readily apparent as the viewpoint moves away from that of the studio camera. This problem can be avoided to a certain extent, if the movement of the viewpoint is restricted to within a small volume around the studio camera, and we use 'billboarding with respect to viewpoint' of the actor model. Hence, the actor model is dynamically orientated so that it remains perpendicular to the optic axis of the viewpoint, instead of the studio camera.

3.2. Shape from Silhouette

In order to overcome the limitations of the planar sprite representation, 3-D information must be incorporated into the model. A fast, scaleable and robust method for reconstruction of 3-D information is shape from silhouette.[7–9]

The silhouette information that is basically needed for this approach is derived by chroma-keying as described in the previous section. Indeed, the sprite model as described above can be interpreted as some kind of shape from silhouette approach using two cameras (one main plus auxiliary camera for the depth).

For better 3-D shape accuracy the number of cameras used must be increased. Experiments show that about 4-6 cameras give a very rough 3-D model of the scene objects, that can be used for enhanced virtual studio applications (corresponding to object class III, Fig. 1). For high-end quality the number of cameras must be further increased and the shape creation must be further enhanced.

The application range of the planar sprite model can be extended even in the case when there is only one main camera available, by using basic assumptions about the object shape: For human beings and some other objects it is a good guess to assume that the volume expansion is maximal in the middle of the object. Further points on the silhouette edge have a big slope, that means the surface normal is almost perpendicular to the optical axis of the camera here. A model that fits to these conditions is to use ellipsoids. These are suited for simple shaped objects and were successfully applied to head and shoulder scenes for modelling persons.[13]

Figure 4. Silhouette mask created by chroma-keying (left), metric of boundary distance (right)

A model that is even simpler to compute and can be used for any shape is a polynomial function of the distance $D(x, y)$ to the 2-D silhouette border:

$$z(x, y) = \frac{s}{2B_{max}}(B_{max} - D(x, y))^2 + H_{max} \quad (1)$$

Where B_{max} is the maximal distance found in the object, H_{max} is the height assigned to this point and s is the slope of the 3-D surface at the silhouette. The distance $D(x, y)$ is implemented using a series of morphological shrink operations. Fig. 4 shows an example.

The creation of a 3-D mesh from this function is then done by a 2-D Delaunay triangulation and setting the z co-ordinate of the triangle vertices by the value found by eq. 1. As a last step a texture map is created from the original coloured input image and assigned to the 3-D description. The final resulting 3-D model is shown in Fig. 7.

The shape creation using a multi-camera system is more complicated: First the camera sensors must be calibrated. That means the internal and external camera parameters must be known. That is done by using a calibration pattern and a calibration procedure to estimate the camera parameter.[12,1] 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. As the last step a texture map is created using the information of the available camera images.

4. RESULTS

This section shows some results for shape creation in the proposed enhanced virtual studio system.

4.1. The 2-D Sprite Actor Model

In Fig. 5, where the virtual and camera viewpoints are coincident, we see the actor partially occluded by an object in the virtual set. This occlusion occurs ‘naturally’ as a part of the rendering process, since the actor model is just another object in the 3-D model. If the virtual viewpoint is slightly displaced from that of the camera, the actor will begin to appear flat. However, with a small amount of displacement, ‘billboarding with respect to viewpoint’ of the actor-plane maintains realism.

In Fig. 6, the virtual viewpoint differs significantly from the camera viewpoint, and it is very clear that the actor is merely a 2-D image placed into the virtual set. Billboarding with respect to viewpoint in this scenario would not resolve the problem, because the viewer expects to see the side of the actor, not the front.

Figure 5. Result of a 2-D sprite actor model; viewpoint coincident with studio camera

Figure 6. Result of a 2-D sprite actor model; viewpoint differs from studio camera

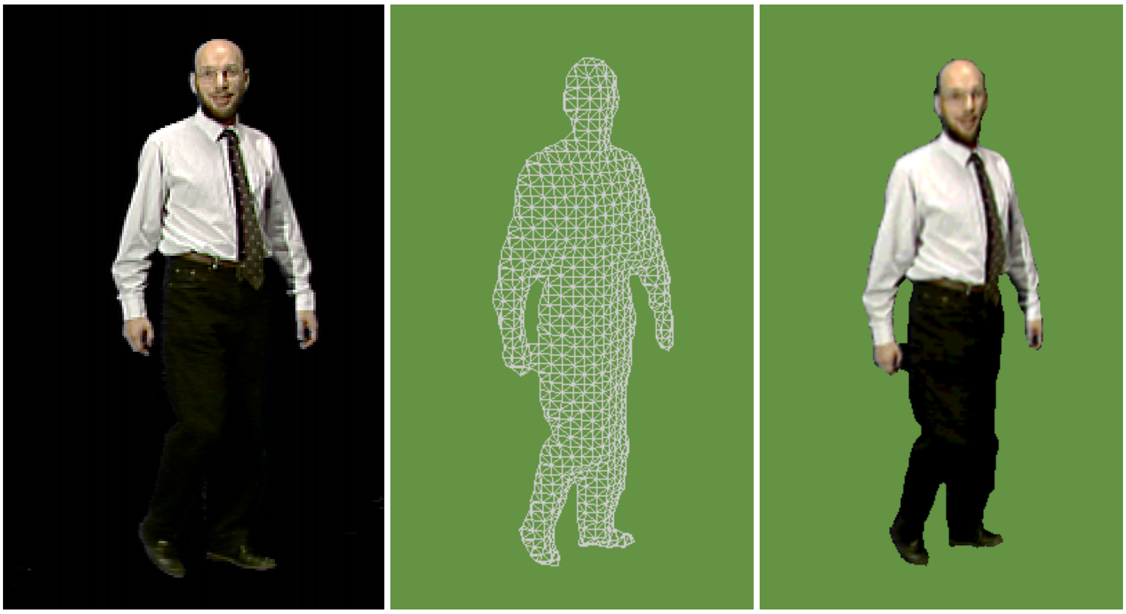

Figure 7. Result of shape from one silhouette (original image (left), created 3-D model after a slight anti-clockwise rotation (middle), rotated textured model (right)

4.2. Shape from one Silhouette

Fig. 7 shows the results of a shape creation by using only one main camera and assumptions about the object geometry, as described in section 3.2. The synthetic image (on the right side) still shows some artefacts, due to the simplified 3-D shape. In contrast with the planar sprite model, as depicted in Fig. 5 and Fig. 6, the appearance when moving the viewpoint away from the original camera is much better.

5. CONCLUSIONS

This contribution gives an overview of applications for virtual production. The analysis of the requirements shows that a high-quality complete 3-D description is not always needed or feasible due to technical limitations. Generally there is a trend to use a simple 3-D representation for real time applications, like the virtual studio. However, our experiments have shown that a high degree of realism can be obtained with a 2-D actor model (class I), based on a video-textured object. That means there is no need for a more complex representation. The trade-off is to constrain the virtual viewpoint to remain close to that of the studio camera used to generate the video-texture.

In applications where the viewpoint cannot be this constrained, a complete (or nearly complete) 3-D description is needed. There might not be the need for a rich detailed description unless, for example, lighting conditions are changed. This is sometimes (not always) the case for special effects.

In order to meet these flexible requirements, an extension of the (classical) virtual studio is proposed. This enhanced virtual studio allows a wider range of possibilities for broadcast applications, as well as new applications, like contributions to interactive 3-D programmes and 3-D animations.

As a suitable technique for 3-D shape reconstruction in the studio environment, shape from silhouette was identified. This method is a direct extension of the current chroma-keying technique, and is robust and scalable. First results show that this method can be used with only two cameras (as shown in section 3.2), incorporating additional assumptions about smoothness of the object surface. For high-accuracy it will be necessary to increase the number of cameras used, and to use additional visual cues like voxel colouring or stereo.

Our future work will involve investigations on estimation of surface properties, i.e. diffuse and specular reflectance. With this information a captured object can be integrated into a scene with a different lighting situation. This feature would be highly desirable in virtual production.

REFERENCES

1. Thomas, G.A., Jin, J., Niblett, T., Urquhatr, "A versatile camera position measurement system for virtual reality TV production", *Proc. of the International Broadcasting Conference*, IEE Conference Publication No. 447, pp. 284-289, September 1997.
2. L. Falkenhagen, "Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continous Surfaces", *Proc. of European Workshop on Combined real and synthetic image processing for broadcast and video production*, 23-24.11.1994, Hamburg, Germany.
3. J.-R. Ohm, et al., "A realtime hardware system for stereoscopic videoconferencing with viewpoint adaptation," in *Signal Processing: Image Communication*, Vol. 14, pp. 147 ff, 1998.
4. M. Ziegler, L. Falkenhagen, R. ter Horst, D. Kalivas, "Evolution of stereoscopic and three-dimensional video," in *Signal Processing: Image Communication*, Vol. 14, pp. 173–194, 1998.
5. P. Kauff, R. Schaefer, O. Scheer, "Tele-Immersion in shared presence conference systems", *Proc. of the Int. Broadcasting Convention (IBC 2000)*, Amsterdam, 8-12 September 2000.
6. M. Price, G.A. Thomas, "3D virtual production and delivery using MPEG-4", *Proc. of the Int. Broadcasting Convention (IBC 2000)*, Amsterdam, 8-12 September 2000.
7. M. Potmesil, "Generating octree models of 3d objects from their silhouettes in a sequence of images",in *Computer Vision, Graphics and Image Processing*, 40:1-20, 1987.
8. R. Szeliski, "Rapid Octree Construction from Image Sequences",in *CVGIP: Image Understanding*, Vol. 58, No. 1, pp. 23-32, July 1993.
9. W. Niem, "Robust and Fast Modelling of 3D Natural Objects from Multiple Views", *SPIE Proceedings*, "Image and Video Processing II", Vol. 2182, 1994, pp. 388-397, San Jose, February 1994.
10. A. Wojdala, "Virtual Studio in 2000: the state of the art", *Virtual Studios and Virtual Production conference*, New York, 17-18 August 2000.
11. R.Calvo, p. Chiwy, C. Girdwood, M. Wells, "ACTS MIRAGE: A success story in virtual reality and 3-D television technology and production", *Proc. of the Int. Broadcasting Convention (IBC 1998)*, Amsterdam, 11-15 Sept. 1998. http://www.itc.co.uk/mirage/
12. 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*, 3 (4) August 1987, pp 323-344.
13. J. Ostermann, "Modelling of 3D moving objects for an analysis-synthesis coder", *Proc. of SPIE/SPSE Symposium on Sensing and Reconstruction of 3D Objects and Scenes*, B. Girod Ed., Proc. SPIE 1260, Santa Clara, California, USA, February 1990.
14. S. Seitz, C. R. Dyer, "Photorealistic Scene Reconstruction by Voxel Scene Reconstruction by Voxel Coloring", *Proc. Computer Vision and Pattern Recognition (CVPR)*, 1997, pp. 1067 1997, pp. 1067- -1073. 1073.
15. U. Kutulakos, S. Seitz, "A Theory of Shape by Space Carving", *Proc Proc. ICCV 99*, 1999.
16. M.J.P.M., Lemmens, "A survey on stere matching techniques" , *Int.Archives of Photogrammetry and Remote Sensing*, vol. 27, Comm. V., pp. 11-23. 1988.
17. A. Hilton, D. Beresford, T. Gentils, R. Smith, W. Sun, "Virtual People: Capturing human models to populate virtual worlds", *IEEE Conf. on Computer Animation*, 1999.
18. S. Weik, J. Wingbermhle, W. Niem, "Automatic Creation of Flexible Antropomorphic Models for 3D Videoconferencing", *Journal of Visualization and Computer Animation*, Vol. 11, pp. 145-154, 2000.
19. L. McMillan, G. Bishop, "Plenoptic Modeling: An Image-Based Rendering System", *Proceedings of SIGGRAPH '95*, Los Angeles, CA, August 6-11, 1995.
20. M. Levoy, P. Hanrahan, "Light Field Rendering", *Proc. of SIGGRAPH '96*.
21. S.J.Gortler, R. Grzeszczuk,R. Szeliski, M. Cohen, "The Lumigraph", *Proc. SIGGRAPH '96*, ACM, 1986, pp. 43-54.