3D MODELLING AND RENDERING OF STUDIO AND SPORT SCENES FOR TV APPLICATIONS

Oliver Grau, Michael Prior-Jones and Graham Thomas

BBC Research & Development, Kingswood Warren, Tadworth, Surrey, UK.
Oliver.Grau | Michael.Prior-Jones | Graham.Thomas@rd.bbc.co.uk

ABSTRACT

The use of 3D modelling and rendering techniques are discussed for applications in TV production. These techniques are very interesting for the creation of special effects and visualisation. In particular the ability to move the viewpoint of the (virtual) camera freely is of interest for the visualisation of sport scenes.

This contribution builds upon our previous work on 3D reconstruction and visualisation techniques based on multi-camera systems for studios. It describes recent work for modelling and rendering of dynamic 3D scenes for studio and sport scenes. In the latter case it describes the practical problems occurring in an outdoor environment.

1. INTRODUCTION

Computer graphics methods and in particular those using 3D models have several applications in TV production. For example they allow the creation of highly realistic special effects. By creating a 3D model of real scene elements full optical interactions, like shadow casting can be established between real and any virtual objects in the scene. Moreover, the viewpoint of the virtual camera can be chosen freely. This technique is also known as free-viewpoint video and is of particular interest for the analysis of sport scenes.

This contribution builds upon our previous work on 3D reconstruction and visualisation techniques based on multi-camera systems for studios [1, 2]. It describes recent work for modelling and rendering of dynamic 3D scenes for studio and sport scenes, as depicted in Fig. 1 + 2. For the latter it describes the practical problems occurring in an outdoor environment.

The approach for the generation of 3D models from dynamic scenes described here uses a time synchronised, calibrated multi-camera system. The studio system [1] is equipped with a chroma-keying facility. Fig. 1 shows some work done in BBC R&D's studio for an experimental production. The cameras in the studio are fixed and can be calibrated with a standard chart calibration method.

In the outdoor environment the situation is more complicated. Even if cameras are mounted in fixed positions

Fig. 1. Scene in the studio.

there are situations where the camera is moving relative to the objects of interest, due to wind or because the entire stand of the stadium is moving under the weight of the audience. Section 2 discusses these problems and describes a dynamic, 'live' calibration that is compensating the changes of the camera parameters during the capture.

Section 4 describes the texture mapping and rendering methods we have implemented.

The contribution finishes with some experimental results and conclusions.

2. CAMERA CALIBRATION

For use in the studio we developed a chart-based calibration method: a 1m x 1m chart showing dots of known size and position is moved in front of the cameras and a set of images is grabbed synchronously. From this data an iterative algorithm computes a set of internal and external camera parameters for all the cameras.

Sport scenes are usually taken in an outdoor environment, like that depicted in Fig. 2. For this environment a chart-based calibration is usually not possible., as the chart would have to be impractically large. As an alternative, we first imvestigated the use of a wand-like calibration object, shown in Fig. 3. By using such an object with a small num-

Fig. 2. A sport scene.

ber of 'features', larger 'features' with a greater separation can be used, which aids their visibility from distant cameras.

The results of the pitch-based methods are compared with those achieved with the calibration wand.

The calibration object depicted in Fig. 3 was used to calibrate 16 SONY DXC 9100P cameras placed around one end of a football pitch. By capturing around 100 images of the object in a series of locations across the visible pitch area, it was possible to use an iterative minimisation procedure to estimate the pose and focal length of each camera.

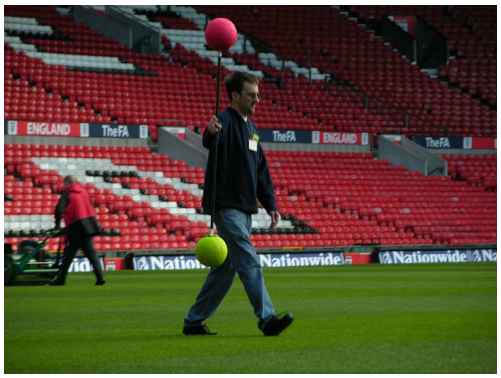

Fig. 3. A calibration object with two balls of different colour.

However, this approach presented several practical problems. The need to capture a specific calibration sequence added to the set-up time, and required careful negotiation with the football ground in order to gain access to the pitch. Furthermore, the reference frame of the camera system was not related to the reference frame of the pitch; it is usually necessary to work in a common reference frame so that the location of the goal and other pitch markings can be reproduced in the dynamic scene models. We also found that there was some movement of the cameras in the period between calibration and the start of the match, partially due to the crowd entering the stadium, causing a small but significant degree of flexing in parts of the stand on which the cameras were mounted. During the game, when the crowd jumped, some cameras also moved.

To overcome these problems, we have developed a dynamic calibration method that uses the positions of the pitch lines in each camera image to calibrate each camera. Line features have been used for camera calibration by previous workers; for example [3] describes a method for registering a camera using lines for outdoor augmented reality. In a method similar to this approach, we first estimate the rough position and orientation of each camera manually. From this estimated pose, we project the estimated position of each pitch line into the image, and identify edge pixels in a region surrounding each estimated line. A straight line is fitted through each set of samples, and outliers further than a given distance from this line (typically 4 pixels) are rejected. The re-fitted lines are then used to compute the 3D camera pose, by minimising the distance from the end points of each fitted line from the projection into the image of the corresponding line in the 3D pitch model (the pitch lines are currently assumed to be of infinite extent). The process is then repeated for every subsequent image. As it is only necessary to process a small neighbourhood around each line, the computational load is relatively small; indeed, it is possible to update the camera pose at full video frame rate using a conventional PC.

One problem is that football pitches are not all of a uniform size, and the size of some pitches can even vary from week to week depending on exactly where the lines are painted. However, this variation generally only applies to the dimensions of the outermost lines; the lines around the goal are usually in well-defined positions. We have therefore extended our calibration method to allow the position of some lines to be fixed only in one dimension (i.e. they lie on the ground plane), so that the unknown dimension can be solved for. Data from all the cameras is used in a single minimisation process, so that a common unknown dimension is determined using all available data.

Figure 4 shows an example of edge points being detected near the positions of predicted lines, and fig. 5 shows the re-projected lines of the virtual pitch. The re-projected lines generally agree with the true pitch lines to better than the width of the painted lines, although one projected line (near the centre of the image) is slightly too high; given the good agreement of the other lines, it is likely that this discrepancy is to an inaccurately-painted line on the real pitch.

3. SEGMENTATION AND 3D RECONSTRUCTION

For the 3D shape reconstruction we use the visual hull method (which is also known as shape-from-silhouette) and a marching cubes algorithm for the generation of a 3D triangle sur-

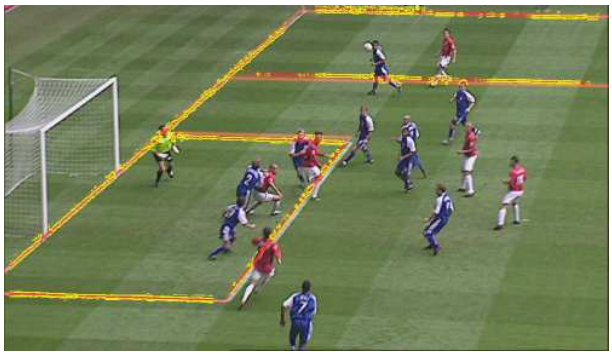

Fig. 4. Edge points (shown in yellow) and fitted straight line segments (in red).

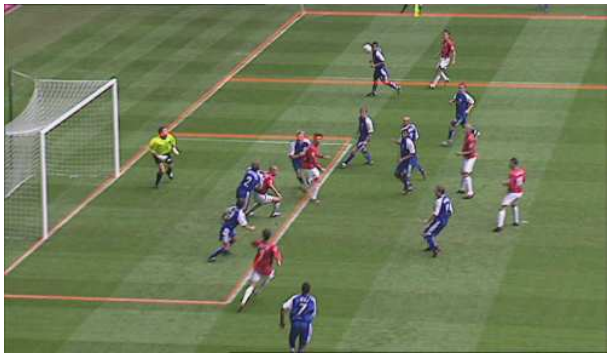

Fig. 5. Re-projected lines (in red).

face approximation. The term visual hull refers to the largest possible object that shares the same silhouettes as the real object when viewed from the same camera angles. It is a convex approximation of the object, and the more camera angles that are used, the closer it is to the shape of the real object. There exist several methods to compute the visual hull, an overview can be found in [4].

Since the volumetric methods 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 [1] and b) super-sampled octree representation [2], which is very robust and used for the work described in this contribution.

The visual hull computation assumes that the silhouettes of the objects of interest are present from several viewpoints. They are computed in the studio using chroma keying techniques. For the outdoor application they are constructed using image segmentation. Two approaches can be applied here: A difference keying and a chroma keying against the green of the grass. Because of the problems found with unstable mounted cameras, chroma keying was used for this work. This techniques quite works well. Problems were found in the the experimental trial with 16 cameras because the images were stored on a server with M-JPEG. The images showed severe compression artefacts that affected the quality of the key.

4. TEXTURE MAPPING AND RENDERING

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

For a high quality rendering in post-production up to three cameras are used and blended. For this purpose the positions of the real cameras are used as nodes in a Delaunay triangulated mesh. The triangle that has an intersection with the line defined by the current position of the virtual camera and the centre of gravity of the virtual scene gives the three cameras that are used for the view interpolation. The blending factors are defined by the barycentric coordinates of the point of intersection in the camera node triangle. Using barycentric coordinates ensures that the blending factors are smooth and continuous on transitions between neighbouring triangles of the camera nodes.

For the rendering we developed an OpenGL-based module that uses the view-dependant texture mapping, as described before. Furthermore an interface to a standard animation package was developed that allows the rendering of high quality images in post-production.

5. RESULTS

Fig. 6 shows the result of the 3D shape reconstruction from a test production in the studio. The scene was captured with 12 Sony DXC 9100P cameras. The voxel resolution was 128x128x128 or approximately 3 cm voxel length. The 3D surface description generated consists of nearly 17,000 triangles. A Gaussian smoothing was applied.

In Fig. 7 an image from a short test production is depicted. A sequence of 3D models (one model per frame) was integrated into a 3D background model and rendered with a standard 3D animation package (Softimage|XSI).

In Fig. 8 a first result of the work done on the sport scene, depicted in Fig. 2 is shown. The 3D models generated with the methods described in chapter 3 have been textured and inserted into a virtual arena. Due to the size of the players in original images the detail in the texture is fairly limited.

6. CONCLUSIONS

The use of 3D modelling and rendering techniques were discussed in this contribution for applications in TV pro-

Fig. 6. Results of the visual hull computation using supersampling rendered as wireframe (left) and in shaded mode (right).

Fig. 7. Rendered scene.

duction in two scenarios: In a studio environment and in an outdoor sport scene. In both scenarios the application of visual hull reconstruction techniques as outlined in section 3 have shown to produce visual results that can be used in the production of special effects or visualisation, like viewing a sport scene from a virtual 'flying camera'.

The work on using these techniques in an outdoor environment is not conclusive yet. More work will be carried out in order to get the dynamic camera calibration more robust and flexible. A solution to integrate images from broadcast cameras, used for the standard TV coverage, is of particular interest. Those cameras are usually operated by cameramen and are following the action. That means the cameras are panned, tilted and zoomed in order to get a better framing of the particular action. These images are very useful for the 3D reconstruction and/or for the texture mapping because they provide more detail than the fixed cameras that operate in wide angle mode. The problem with zoomed in cameras is that not always enough line features from the pitch are visible. Therefore, a solution working directly on the silhouettes for camera calibration is envisaged.

Fig. 8. Sport scene integrated into virtual arena.

Further work will be carried out in the texturing and rendering of the 3D models, since there are cases where the simple blending as outline in section 4 is not sufficient, e.g. when cameras are showing only parts of the scene.

7. ACKNOWLEDGEMENTS

This work has been funded by the EU projects IST-MATRIS and IST-ORIGAMI. The background model in Fig. 7 was provided by CAU University of Kiel.

8. REFERENCES

[1] 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.

[2] O. Grau, "3d sequence generation from multiple cameras," in *Proc. of IEEE, International workshop on multimedia signal processing 2004*, (Siena, Italy), September 2004.

[3] B. Jiang, U. Neumann, and S. You, "A robust hybrid tracking system for outdoor augmented reality," in *Proceedings of 2004 IEEE Conference on Virtual Reality*, pp. 3–10, 2004.

[4] C. R. Dyer, "Volumetric scene reconstruction from multiple views," in *Foundations of Image Understanding*, L. S. Davis, ed., pp. 469–489, Kluwer, Boston, 2001.