THE QUEEN'S AWARDS FOR ENTERPRISE 2001

BBC

R&D White Paper

WHP 102

December 2004

3D sequence generation from multiple cameras

O. Grau

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 102

3D sequence generation from multiple cameras

Abstract

An approach for the generation of 3D surface model sequences from a set of synchronised, calibrated cameras is presented. The method is based on the computation of the visual-hull.

An analysis classifies typical error sources for artefacts in the reconstructed 3D models into occlusion, approximation and sampling errors. For the synthesis of new images for TV- and film-applications the roughness of the reconstructed 3D models has shown to be very disturbing. In order to reduce this roughness two methods are used sequentially: A new super-sampling approach that reduces the sampling error, and a Gaussian surface smoothing. The application of the modelling results from a 12 camera studio system is shown for the generation of virtual views of moving actors in standard post-production packages.

This document was originally published in Proc. of IEEE, International workshop on multimedia signal processing 2004, (Siena, Italy), September 2004

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

3D sequence generation from multiple cameras

Oliver Grau
BBC Research & Development
Kingswood Warren
Tadworth, Surrey, KT20 6NP, UK
Email: oliver.grau@rd.bbc.co.uk

***Abstract*— An approach for the generation of 3D surface model sequences from a set of synchronised, calibrated cameras is presented. The method is based on the computation of the visual-hull. An analysis classifies typical error sources for artefacts in the reconstructed 3D models into occlusion, approximation and sampling errors. For the synthesis of new images for TV- and film-applications the roughness of the reconstructed 3D models has shown to be very disturbing. In order to reduce this roughness two methods are used sequentially: A new super-sampling approach that reduces the sampling error, and a Gaussian surface smoothing. The application of the modelling results from a 12 camera studio system is shown for the generation of virtual views of moving actors in standard post-production packages.**

I. Introduction

THE use of 3D models opens many possibilities for special effects in TV- and film-productions and is increasingly used. A closer look reveals that these productions mainly make use of manually created content by post-production houses or scanned static 3D objects. The actual dynamic 'live-data' is mainly used as 2D layered data from one camera at a time. Equally current research focuses mainly on 3D modelling of static objects.

This contribution describes an approach to generate 3D information of dynamic scenes, in particular of actors or presenters in a studio environment as depicted in Fig. 1. The approach currently requires a segmentation of the scene into foreground and background by chroma-keying. The generated 3D data is used in post-production for special effects or in extended virtual studios [1].

A. Related work

The 3D-capture of dynamic scenes requires a robust and highly automatic approach. The computation of the visual hull from 2D silhouettes has been shown to be quite robust, fast and suited for many practical object classes [2], [1], [3]. The computation of the visual hull, also known as shape-from-silhouette is equivalent to an intersection of the back-projected 2D silhouettes (visual cones) in 3D. A number of approaches to compute this intersection have been suggested in the literature that differ in the underlying data structure and the processing [4], [5], [6], [2], [7], [8]. Although the method has been applied to the computation of sequences, relatively little effort has been made to analyse and address the problems specific to sequence processing.

Fig. 1. Scene in the studio from a demo production

The approach described in this article uses a time synchronised, calibrated multi-camera system equipped with chroma-keying facility [9]. The visual hull is then computed using a volumetric octree representation and hierarchical processing [5]. A surface description is then computed using the marching cubes algorithm [10]. As a novel approach we apply super-sampling to this method and show how this is reducing the quantisation noise during the surface generation.

The paper is organised as follows: The next section analyses the possible artefacts of 3D models generated with shape-from-silhouette methods with special focus on dynamic objects. Section III introduces our processing using a new super-sampling approach. The paper finishes with experimental results in section IV and conclusions.

II. Analysis of artefacts in dynamic visual hull reconstructions

The visual hulls of dynamic objects can show a number of artefacts that are caused mainly by the following three error sources:

- *Occlusion errors* include the fundamental limitation of the visual hull to fail on concavities. The significance of this error class depends strongly on the shape of the modelled objects. Further the number and positions of the cameras has an important influence on the result of the shape generation, i.e. the fewer the number of cameras, the more likely artefacts due to occlusions will become visible.

- *Approximation errors* are due to a low number of cameras. The hatched areas in Fig. 2 indicate the approximation error of a reconstructed object from only two cameras. The reconstruction is always bigger than the actual object. A quite disturbing effect of the approximation error in the context of time sequences are *moving edges*: When an object moves the edges formed by two adjacent camera views move relatively over the object which is visibly quite disturbing in a synthesised image of the reconstruction.
- *Sampling errors* occur in the volumetric domain by choosing an inappropriate resolution or in the conversion from volumetric to surface description. To choose the resolution one argument is that it is not useful to use a volumetric resolution in which the footprint [1] of a voxel [2] is smaller than an actual pixel. In practice the resolution is significantly lower than this in order to save processing time and to keep the complexity of the reconstructed models low.

The sampling or quantisation error during conversion from the volumetric to the surface description was analysed in [7], [11]. The problem is that conventional implementations of a visual hull reconstruction use binary voxels. The surface generator (usually a marching cubes algorithm) introduces quantisation noise into the surface description. In [7] we made a suggestion to use line-segments with non-quantised length in order to avoid this problem. The next section introduces super-sampling of volumetric data that delivers a similar reconstruction quality.

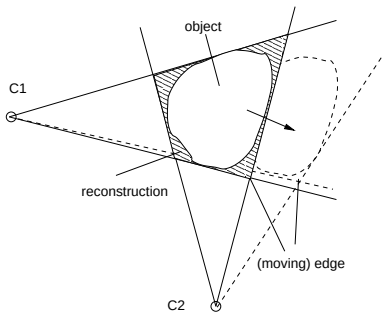

Fig. 2. Visual hull reconstruction from two cameras

Fig. 3 illustrates effects of the error sources described. The reconstruction was done using only six cameras and an octree-based data structure with hierarchical processing. The voxel resolution is 128x128x128. In both views of Fig. 3 the problems due to occlusions can be seen: There are remaining bits on the feet and the separation of the two persons is quite poor. The approximation problems due to the low number of cameras is clearly visible in the top view of Fig. 3: In particular the person on the right shows a 'blocky' shape.

[1]projection into the image plane
[2]voxel = volume element

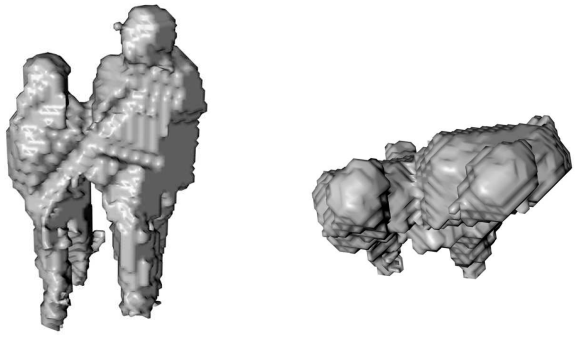

Fig. 3. Result using 6 cameras and voxel-based 3D reconstruction (128x128x128) front view (left) and top view (right)

Both problems can be overcome by using more cameras: By upgrading the studio system from 6 to 12, shape quality is improved, as can be seen in Fig. 4.

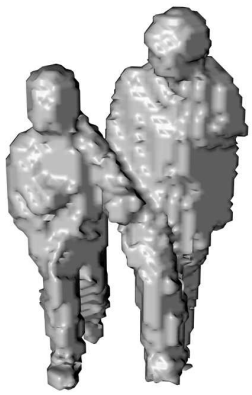

Fig. 4. Result using 12 cameras and voxel-based 3D reconstruction (128x128x128)

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

Another problem visible in Fig. 4 is the fact that the surface looks 'voxelised', meaning that the voxel structure is visible due to the sampling error during the surface generation as mentioned above.

III. An improved shape generation approach

This section introduces a new variant of the octree-based volumetric reconstruction that reduces the sampling error.

An octree is a hierarchical data structure that divides the volume on each level into 8 octants. Therefore the resolution of an octree is determined by its maximum level L_{max}:

$$N_x = N_y = N_z = 2^{L_{max}} \tag{1}$$

On each hierarchy level L the reconstruction algorithm is testing whether an octant is foreground or not. This test is done by projecting the octant into all silhouette images and checking if the footprint of the octant is intersecting with any foreground pixels of the silhouette. If the intersection is zero that means that it is completely background and the octant can be set to $false$ and no further testing is required. In the case there is intersection in all silhouette images the octant will be further sub-divided until the maximum level L_{max} is reached. In this conventional octree-based reconstruction the value of a voxel can be either $true$ or $false$ which gives the known quantisation error in the surface description from the marching cubes algorithm.

We suggest here to perform a further super-sampling of the octree. That means we further subdivide an octant O_i to the maximum level L_{max} by L_{ss} levels and assign to it the sum of $true$ sub-voxels:

$$V_{O_i} = \sum_{j=1}^{L_{ss}} \sum_{k=0}^{7} V_{O_i}(j,k) \tag{2}$$

The marching cube is then used to compute the iso-surface with a threshold that is usually half the number of sub-voxels, i.e.:

$$C_{isothreshold} = 2^{L_{ss}}/2 \tag{3}$$

A. Gaussian Smoothing

The reconstruction using super-sampling still has a rough surface due to noise in the silhouette images and shows *moving edges* due to approximation errors. This is particularly disturbing in a sequence of 3D models because a significant flickering can be observed.

We therefore apply a simple Gaussian smoothing filter on the surface by replacing the coordinates of a vertex V_i by the average of its K neighbouring vertices:

$$V_i = (X_i, Y_i, Z_i)^T = \frac{1}{K} \sum_{j}^{K} V_j \tag{4}$$

B. Complexity reduction

The 3D models in the previous examples, like in Fig. 4 typically contain around 20,000 triangles which generates a large amount of data when processed in a sequence. In order to reduce the complexity we are optionally using a polygon reduction method as described in [12].

IV. Results

This section shows a few experimental results taken in a studio environment equipped with chroma-keying facility [9]. The images where grabbed by a calibrated, synchronised multi-camera system using 12 SONY DXC-9100P cameras with 704 x 576 resolution captured at 25 fps progressive.

Fig. 5 shows a reconstruction using the octree reconstruction using one additional super-sampling level, i.e. $L_{ss} = 1$.

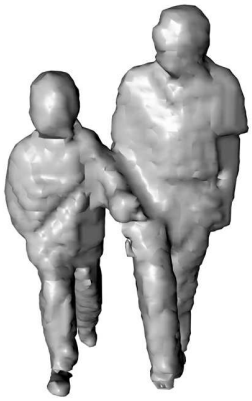

Fig. 5. Result using voxel-based 3D reconstruction with super-sampling

In Fig. 6 and Fig. 7 the same object after applying Gaussian smoothing is depicted.

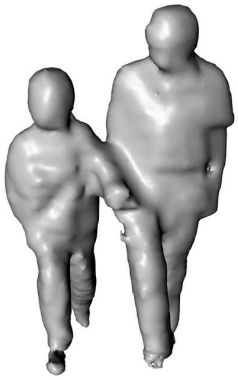

Fig. 6. Result using voxel-based 3D reconstruction with super-sampling and Gaussian smoothing

Fig. 8 shows the resulting surface model after complexity reduction to 3000 triangles. The desired number of triangles depends on the application requirements and can be varied.

The techniques described in this article have been tested in a recent demo production. A sequence of actions was captured in the studio (see Fig. 1) and processed later. The resulting 3D models were then integrated into background models created

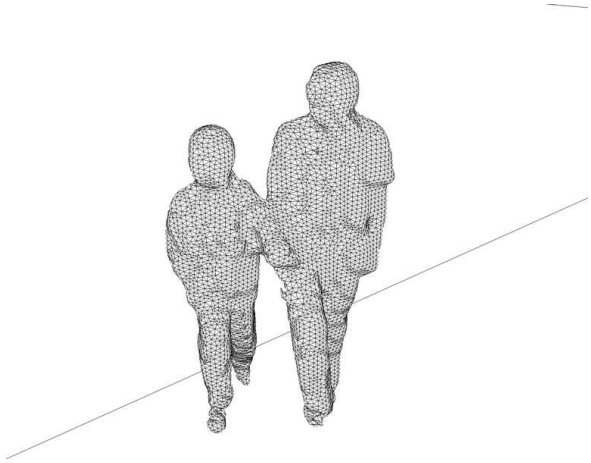

Fig. 7. Result using voxel-based 3D reconstruction with super-sampling and Gaussian smoothing (wireframe)

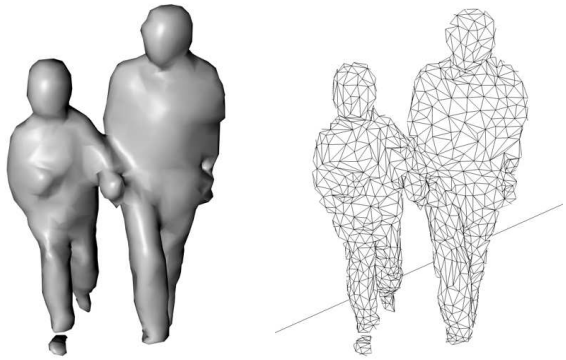

Fig. 8. Result using voxel-based 3D reconstruction after polygon reduction (wireframe)

by Kiel University (CAU) within the IST-ORIGAMI project. Fig. 9 shows one frame of a 10 s clip.

V. Conclusions

An approach for the generation of sequences of 3D surface models from a set of synchronised, calibrated cameras was presented. An analysis was made which classified typical error sources for artefacts in the reconstructed 3D models into occlusion, approximation and sampling errors.

For the synthesis of new images for TV- and film-applications the roughness of the reconstructed 3D models using conventional visual hull computation methods was considered to be visually disturbing. In order to reduce this roughness two methods are used sequentially: A new super-sampling is introduced that reduces the sampling error and a Gaussian smoothing is used to further reduce the surface roughness.

The results in the previous section show the effect of this processing chain. The quality of the 3D actor shape is good enough to change the lighting slightly and to integrate the models into a different virtual environment as depicted in Fig. 9 so that it can be used in the targeted applications.

Fig. 9. Resulting frame of a video sequence with 3D model integrated into virtual background

References

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

[2] 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. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000, pp. 369–374. [Online]. Available: citeseer.nj.nec.com/article/matusik00imagebased.html

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

[4] 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.

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

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

[7] 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)*, London, UK, April 2003, pp. 395–404.

[8] T. Matsuyama, X. Wu, T. Takai, and T. Wada, "Real-time dynamic 3-d object shape reconstruction and high-fi delity texture mapping for 3-d video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 357–369, March 2004.

[9] 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*, vol. 14, no. 3, pp. 370–380, March 2004.

[10] 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*. ACM Press, 1987, pp. 163–169.

[11] O. Grau, "A studio production system for dynamic 3d content," in *Proc. of SPIE, Visual Communications and Image Processing 2003*, vol. 5150, Lugano, Switzerland, July 2003.

[12] M. Garland and P. Heckbert, "Simplifying surfaces with color and texture using quadric error metrics," in *Proc. of IEEE Visualization 98*, 1998.