Automatic Modelling of 3D Natural Objects from Multiple Views

Wolfgang Niem, Ralf Buschmann
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstr. 9A, D–30167 Hannover, Germany
email: niem@tnt.uni–hannover.de, buschman@tnt.uni–hannover.de

Abstract

An algorithm for the fast automatic construction of a 3D model of any real object using images from multiple views is presented. The images are taken from a real object rotating in front of a stationary calibrated CCD TV camera. The presented algorithm generates the object shape in a first step. For that purpose an effective implementation of the method of occluding contours is used to obtain a convex volume model of the object. This model is refined in order to detect shape concavities by using additional depth information from disparity estimation and finally approximated by a triangle mesh. In a second step the texture is estimated from the image sequence and projected onto the surface model to obtain natural looking models. Results with real image sequences have confirmed the suitability of the developed algorithm even for the modelling of real objects with highly detailed and complex surfaces.

1 Introduction

The objective of this work is the development and refinement of algorithms for modelling arbitrary shaped 3D objects using multiple perspective views from different viewpoints. Fields for applications are computer animation used in TV production, architecture simulation, as well as driving and flight simulators. For all these applications the generation of the virtual scene can be facilitated by providing a large library of model objects. It is important that the acquisition of the model objects can be done with simple equipment and in reasonable time.

One way of constructing a virtual 3D–model is the evaluation of multiple views. For a turntable with the real object rotating in front of a stationary calibrated camera Busch [2] describes a technique which can be divided into three steps. In a first step a rough volume model is estimated using the silhouettes of the real object applying the method of occluding contours [3],[8],[9],[10],[11]. The surface of the volume model is approximated by a triangular mesh in a second step and in a third step the real object texture is projected onto this surface model to obtain a natural looking model.

One drawback of this technique is the volume modelling algorithm which is only suited for simple–shaped convex objects and uses an inefficient volume representation. Further it shows deterioration of edges in the 3D model and the simple method for texturing of the surface model leads to visible defects of the texture.

In order to overcome the described drawbacks three new approaches are investigated.

The first approach exploits a new volume representation during volume modelling. Further an algorithm for the detection of object concavities which can be integrated into the volume model is developed. The second approach aims the representation of the volume model by a triangular mesh which is realized by considering 3D–edge information. A method for the integration of texture information from different images is proposed in a third approach in order to reduce texture distortion.

Section 2 describes the system calibration and the camera model used for the modelling process. Section 3 deals with volume modelling. Section 4 explains the surface modelling method. Section 5 describes the texture estimation technique. Section 6 presents and discusses results obtained with the developed new algorithm.

2 System Calibration and Camera Model

The environment used for this work consists of a stationary CCD TV camera in front of a turntable, which can be rotated in controlled steps. The background is of uniform colour in order to facilitate the segmentation of the real object against the background. Before the acquisition of a real object on the turntable can be started, the system has to be calibrated in order to obtain position, orientation and focal length of the camera. For that purpose, a precisely known test pattern consisting of black circles on a white background is placed on the turntable as shown in Fig. 1. The detection of the corresponding ellipses in the image taken by the CCD camera and the subsequent estimation of all parameters describing the acquisition process is performed using a calibration method like the one proposed by Tsai [13].

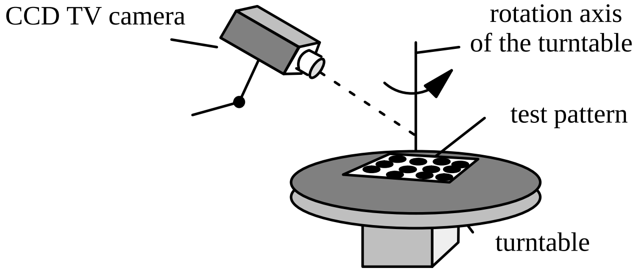


Fig. 1 Arrangement for system calibration

For the processing of the input images, the real camera must be represented by a mathematical model. A camera model, which is suited for 3D–modelling applications, is the one introduced by Yakimovski and Cunningham [14]. It assumes the camera to be geometrically linear, an assumption which is reasonable, considering the linear sensor array and the high quality of the used lens. Thus, it is possible to use the laws of central projection for the object acquisition.

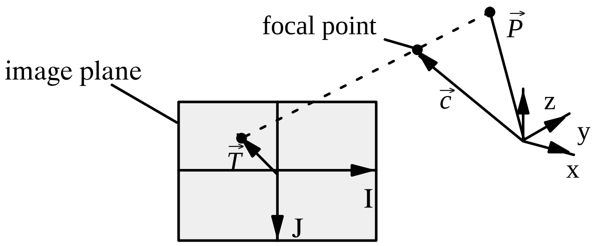


Fig. 2 Projection of a real–world point $\vec{P}$ into an image point $\vec{T} = (I,J)$

Denoting

$\vec{c} = (c_x, c_y, c_z)^T$: vector to the focal center
$\vec{a} = (a_x, a_y, a_z)^T$: unit vector in the direction of the optical axis
$\vec{h} = (h_x, h_y, h_z)^T$: unit vector parallel to the horizontal axis of the image plane
$\vec{v} = (v_x, v_y, v_z)^T$: unit vector parallel to the vertical axis of the image plane
f: : focal length of the camera

the projection (Fig. 2) of a point $\vec{P}$ into the image point $\vec{T} = (I,J)$ can be performed as

$$I = \frac{(\vec{P} - \vec{c}) \cdot \vec{h}}{(\vec{P} - \vec{c}) \cdot \vec{a}} \cdot f \quad ; \qquad J = \frac{(\vec{P} - \vec{c}) \cdot \vec{v}}{(\vec{P} - \vec{c}) \cdot \vec{a}} \cdot f \tag{1}$$

3 Volume Modelling

The developed algorithm for the construction of a volume model consists of three steps. In a first step a bounding volume is estimated using the object silhouettes; in a second step additional depth information from different viewpoints is fused and in a third step the fused depth information is integrated into the volume model to reduce modelling errors (Fig. 3).

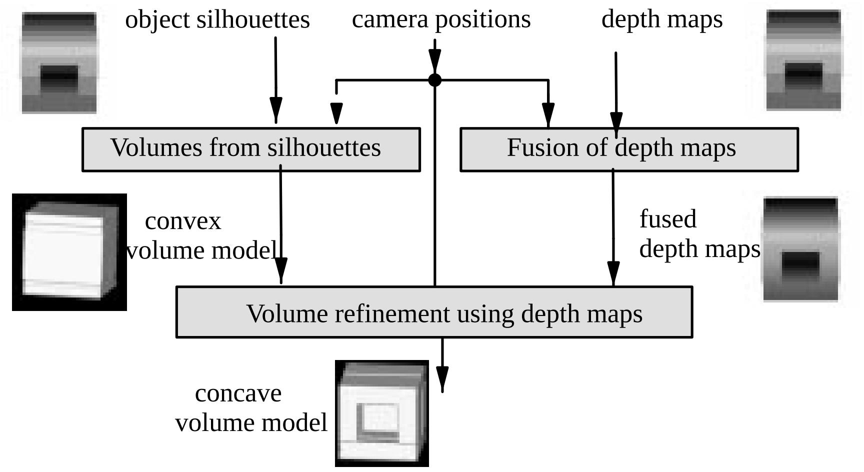


Fig. 3 Block diagram of the developed algorithm for volume modeling

3.1 Volumes from silhouettes

An approach for the construction of volume models using multiple views was first described by Martin and Aggarwal [7]. They introduced the method of occluding contours which uses the object silhouettes and the related camera parameters to construct a volume model of the real object. A key point in performing this method is a proper volume representation, characterized by low complexity and suitability for a fast computation of volume models.

In this work, the volume is decomposed into pillar–like volumes (pillars) which are built of elementary volume cubes of the finest resolution. Each of those pillars is completely described by the position of the center points of the cubes on the top and the bottom of the pillar as shown in Fig. 4. The complexity of this representation is proportional to the object surface area.

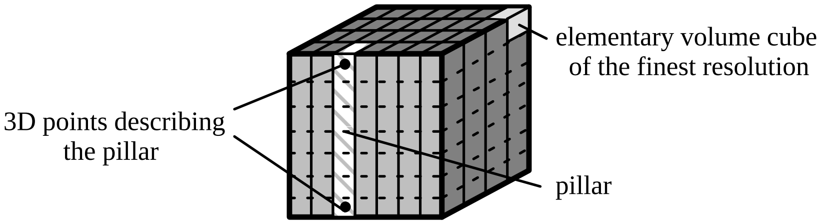


Fig. 4 Volume representation using pillar–like volume elements (pillars)

The volume modelling algorithm starts with the definition of an initial volume cube that surely contains the object. The algorithm works as follows:

1. Segmentation of the input image to get the object silhouette
2. Projection of the 3D points describing the pillar into the image plane with the silhouette (Fig. 5a)
3. Connecting the resulting image points with a 2D line (Fig. 5b)
4. Subdividing and reducing the 2D line into new line segments by eliminating the pixels outside the silhouette (Fig. 5c)
5. Subdividing and reducing the pillar corresponding to the pixels describing the new 2D line segments (Fig.5d)

Applying those steps consecutively to each pillar, the approximation of the volume model represented by a set of pillars will improve for each processed silhouette.

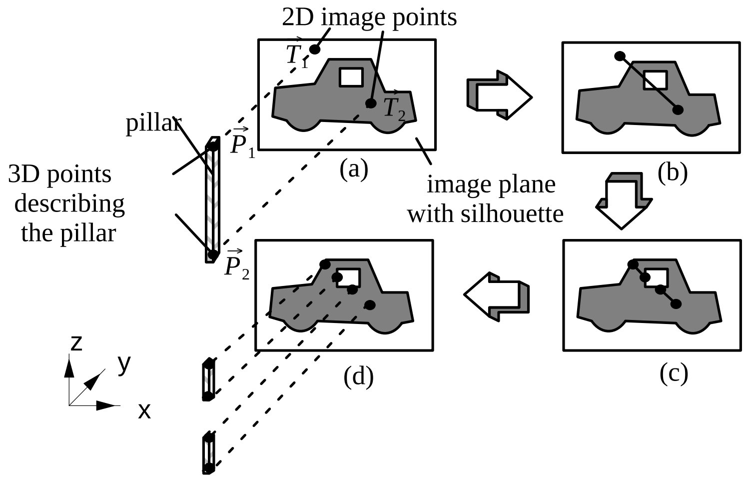


Fig. 5 Volume construction using pillars

(a) Projection of the 3D points $\vec{P}_1$, $\vec{P}_2$ into the image plane with the silhouette

(b) Connecting the resulting 2D image points with a 2D line

(c) Subdividing and reducing the 2D line into new line segments by eliminating the pixels outside the silhouette

(d) Subdividing and reducing the pillar corresponding to the pixels describing the new 2D line segments

The application of this algorithm can lead to the following two modelling errors: a) errors due to the restriction of the algorithm to convex objects and b) errors caused by the finite number of views used for the volume modelling. In Fig. 6 these two errors are shown exemplary**.**

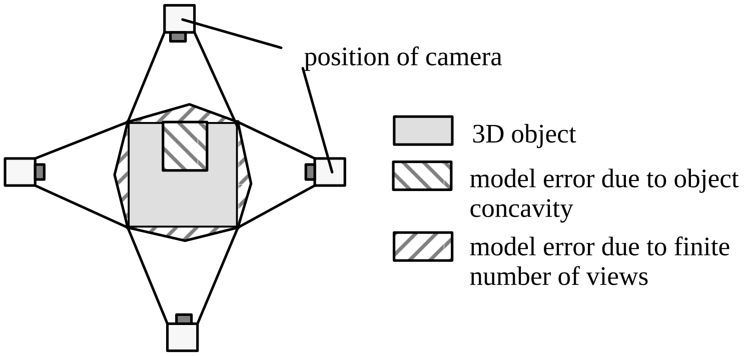


Fig. 6 : Two dimensional description of the resulting volume model

3.2 Integration of depth information

To overcome the principal inability of the above described volume modelling technique to model objects with concave shape additional information about the distance of the objects surface to the image plane is needed. This distance information, in the following called depth information, is stored in depth maps where a depth value is assigned to each pel. To obtain the depth maps active methods like laser scanner or structured light as well as passive methods like several disparity estimation techniques exist. Here the application of disparity estimation techniques is possible as the camera parameter, the 3D geometry between camera and turntable, and the rotation angle of the turntable between two successive images are exactly known. Thus two successive images can be viewed as one stereo image pair, and after rectification standard disparity estimation techniques can be used. As for active methods always additional equipment is needed, in this work a disparity estimation technique is used which is described in detail in [4]. This technique does not only estimate disparity values but also assigns a reliability value to each disparity value.

In general the depth maps estimated by active methods or by disparity estimation are inaccurate due to measurement errors and due to their restricted amplitude resolution. In the following a method is presented which describes how the accuracy of depth maps and their reliability can be increased by fusing depth maps from different viewpoints.

In a first step the depth of a surface element of the object is determined by considering the different depth values estimated from neighbouring viewpoints from which the surface element can be seen. Therefor at first a reference depth map is chosen. Then the 3D point represented by the assigned depth value z_1 of a pel $\vec{T}(x_1, y_1)$ of a neighbouring depth map is transformed into the reference depth map resulting in a 2D point $\vec{T}(x_0, y_0)$ and the depth value z_0' (Fig. 7).

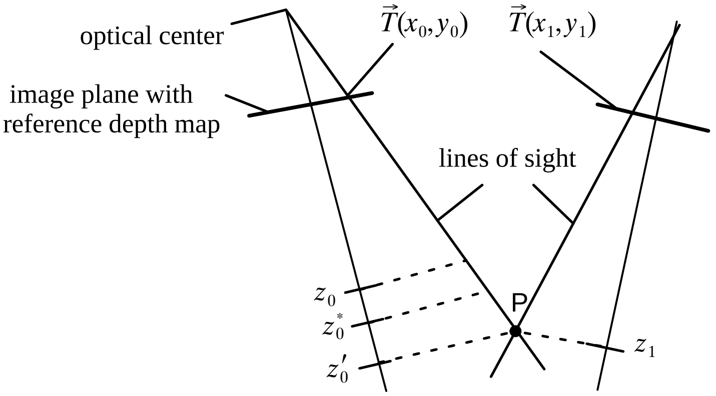


Fig. 7 Fusion of depth maps

P represented by the depth value z_1 of a pel $\vec{T}(x_1, y_1)$ of a depth map is transformed into the reference depth map resulting in $\vec{T}(x_0, y_0)$ and the depth value z_0' . The mean depth value of z_0 and z_0' denoted z_0^* is assigned to $\vec{T}(x_0, y_0)$.

If the difference between the assigned depth value of $\vec{T}(x_0, y_0)$ in the reference map z_0 and z_0' is below a predefined threshold, the mean depth value of z_0 and z_0' denoted z_0^* is calculated and assigned to $\vec{T}(x_0, y_0)$ considering their assigned reliability value.

The threshold avoids wrong correspondences for object surface elements which are visible from one viewpoint, but which are hidden in the neighbouring viewpoint. This procedure is repeated for each pel of each neighbouring depth map. In this way the depth measures of neighbouring depth maps are exploited to result in a more reliable and more accurate fused depth map.

In a second step an interpolation of the depth values assigned to neighbouring surface elements is performed, as neighbouring surface elements are likely to have similar depth values. Using physical surface models (membrane and plate) [12] a method for the smoothing of depth maps has been developed. In order to avoid errors due to smoothing at object edges, the edges are estimated during the smoothing process and considered as discontinuities in the physical models. The spatial interpolation process has thus been defined as a non–convex optimization problem. The optimization is performed using the GNC (Graduated Non–Convexity) algorithm, which has originally been developed for the reconstruction of luminance images [1]. Simulations have shown that especially the physical plate model in conjunction with the GNC algorithm can be recommended for the spatial interpolation process.

To enable the modelling of concavities the fused depth maps are used to refine the volume model. For each pel of each viewpoint the voxels are eliminated along the line of sight up to the assigned depth values in the fused depth maps. Modelling errors may now occur due to inaccurate depth information (too large depth values) and due to the limited number of viewpoints. The limited number of viewpoints leads to errors in areas of hidden surfaces (Fig. 8).

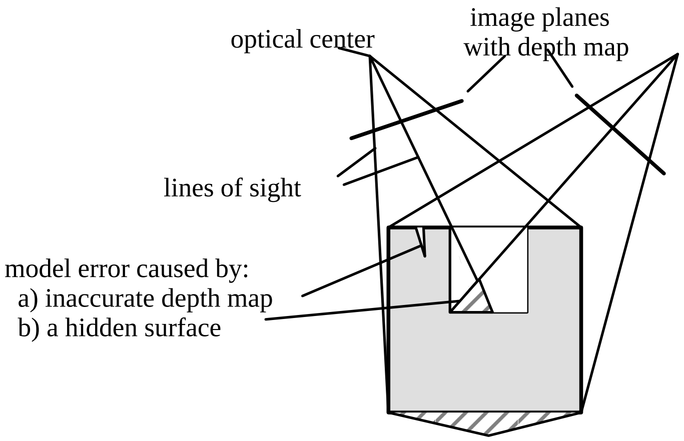


Fig. 8 Modelling of concavities by the use of depth maps

In the visible surface area the concavity is carved out by eliminating the volume elements along the line of sight of each picture element up to its assigned depth measure. Model errors caused by inaccurate depth maps and hidden surface are shown.

4 Surface Modelling

The synthesis algorithms used in computer animation work with surface models. Thus the volume representation must be transformed into a surface model. For that purpose the surface of the volume model is approximated by a triangle mesh. A mesh growing algorithm is used which allows a local adaptation of the volume model surface by adapting the size of each triangle patch according to a tolerable approximation error. This results in smaller triangles in surface regions with high curvature and larger triangles in surface regions with low curvature.

Starting with a single triangle placed on the volume model surface, or with a set of triangles located at 3D edges of the volume model, further triangles are constructed at each open edge until the whole volume model is covered with a mesh (Fig. 9).

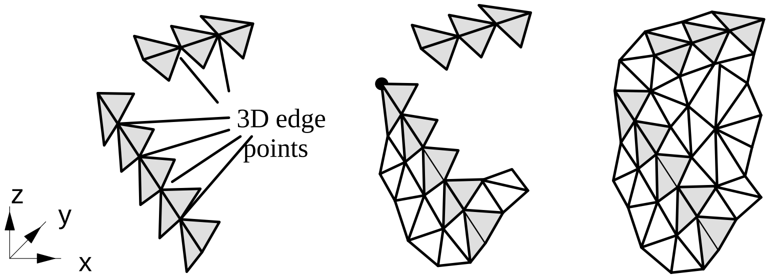


Fig. 9 Approximating the volume model by a triangle mesh

For the construction of the triangle mesh the following quality criteria have to be taken into consideration. These criteria can be divided into two groups. The criteria in the first group control the properties of each new single triangle, whereas the criteria in the second group evaluate the local neighbourhood of the triangle to form a regular mesh.

Criteria for single triangles are:

1. The distance between the triangle surface and the assigned surface voxels of the volume model must be within a tolerance to fulfill a given approximation quality. This leads to a generation of large triangles in planar areas and to small triangles in areas with large curvature.
2. The triangles should be equilateral and all angles should be larger than a given minimum angle. This avoids slim triangles which would lead to a direction dependant formability of the mesh.
3. The maximum size of the triangles is restricted to provide a high formability even in planar regions of the object surface.

Criteria for the mesh:

4. The distance of a new vertex to all edges of existing triangles should be larger than a given minimum triangle height. This avoids the need to fill the gap within the triangles with a slim triangle.
5. A gap between neighbouring triangles with an angle less than 80° will be closed directly, if allowed by the other criteria. This improves the symmetry of the mesh.

6. A gap between neighbouring triangles with an angle between 80° and 120° will be closed by symmetrical insertion of a new vertex. This is not really needed, but like criterion 5. improves the symmetry of the mesh forcing all angles to be between 40° and 80°.
7. If a new triangle overlaps an existing triangle, the new vertex will be fixed to the nearest vertex of the existing triangle, in order to avoid further overlapping. This is also how the mesh is closed. The recognition of overlapping triangles is very complex.
8. If no new triangle can be added to an existing edge, then the surface of the object is very uneven in this area and can not be approximated by a triangle with the given edge length. The edge has to be divided to enable the building of smaller triangles.

5 Texturing

The geometric model of the real object is not sufficient for applications in the field of computer animation. For natural looking 3D models the model texture is of great importance. To meet this goal a texturing algorithm has been developed which estimates the texture for each surface triangle from the input image sequence.

The principle to bind texture information to a single surface triangle is explained in Fig. 10. By using the cahv–camera parameters the vertices of a surface triangle are projected into the camera plane with the original image. The clipped rectangular image part containing the projected triangle is defined as texture map.

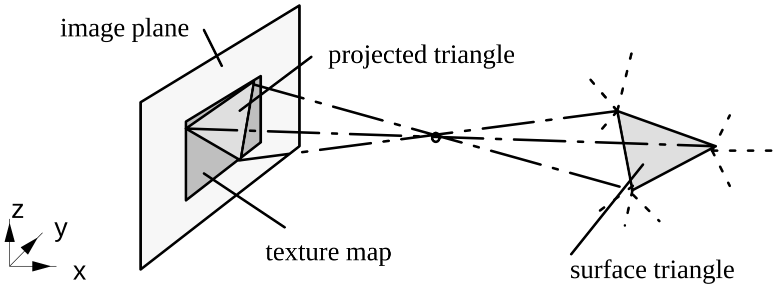
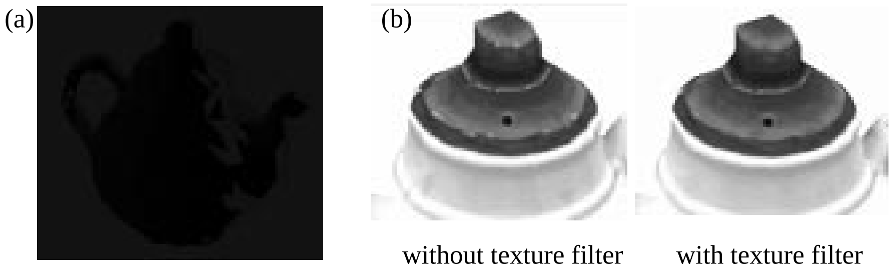


Fig. 10 Binding a texture map to a surface triangle

The quality of the final texture depends on illumination and the resolution of the CCD camera. Furthermore the correct binding of the texture to the 3D surface model is influenced by the accuracy of camera position relative to the object and by shape errors. All these influencing factors have to be taken into account in the texturing process in order to obtain a model with low distortion of texture.

The proposed texturing method is divided into three steps:

1. *Grouping triangles to surface regions textured from a common image***.**(Fig. 11a) As the texture at boundaries between surface regions textured from different images is probably distorted, those regions should be as homogeneous as possible in order to reduce the total length of boundaries. Furthermore, the assignment of a triangle to a surface region depends on its texture resolution, which is defined as the ratio of pixel elements per surface unit and should be tolerable within a surface region.
2. *Local texture filtering of the boundaries between the surface regions*. (Fig. 11b) This is achieved by a newly developed filter which blends the texture between neighbouring triangles of different surface regions. This results in a local blurring effect, which is less conspicuous then the blocking effect occurring without filtering.
3. *Synthesis of texture for surface triangles not visible in any image.* For that purpose a filter has been developed which uses the texture from visible triangles.



without texture filter with texture filter

Fig. 11 (a) Grouping triangles to surface regions (b) effect of the texture filter

6 Results

The described algorithm was tested on a number of real 3D objects. For that purpose, the real object was placed on a turntable and the turntable was rotated in 10^0 steps. For each of the 36 views, a 720 x 576 image was taken from the real object with a stationary calibrated CCD camera. The rotation angle accuracy of the turntable was about 0.05^0. The processing of the volume was performed using a resolution in space of 160^3 unit cubes for the bounding box of each real object.

Input sequences were taken not just from simply shaped objects but also from natural objects with complex surfaces.

Fig. 12–14 show exemplary synthezised projections of the automatic generated models from a dinosaur, asterix and a wooden car. The results show a quality which is sufficient for computer animation in case of the complex convex models asterix and dinosaur. In case of the concave model of the wooden car the quality is lower which can be traced back to errors in the depth maps.

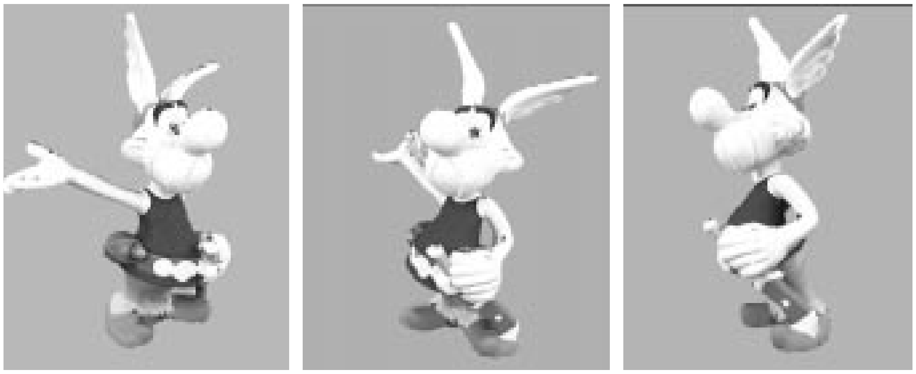


Fig. 12 Synthetic generated projections of the model ”asterix”

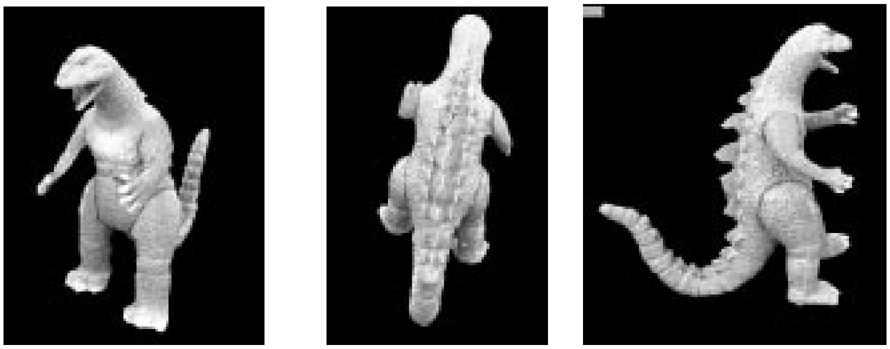


Fig. 13 Synthetic generated projections of the model ”dinosaur”

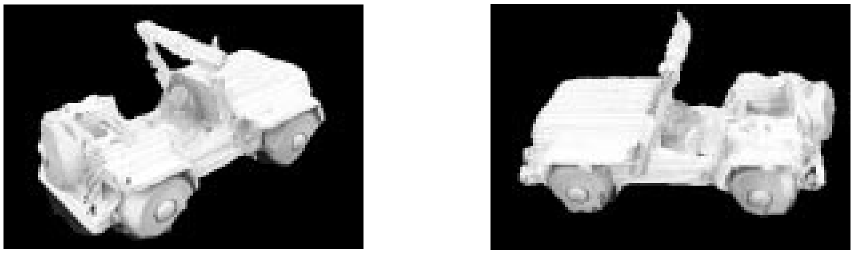


Fig. 14 Synthetic generated projections of the model ”wooden car”

7 Conclusions

A complete process chain for the automatic modelling of 3D natural objects from multiple views is presented.

Based on a new volume representation the full resolution accuracy of the object silhouettes during volume modelling is exploited. The new volume representation furthermore allows a fast and simple realisation of the method of occluding contours. In order to enable the modelling of concave objects, additional depth information obtained by disparity estimation is integrated into the convex volume model.

The representation of the volume model by a triangular mesh is realized by considering 3D–edge information. Vertices of the triangular mesh can explicitly be set to 3D–edge coordinates in the initialization phase of the algorithm, which leads to reduced deterioration of edges in the final surface model.

For the texturing of the surface model a new method using all available images is proposed. Neighbouring triangles describing the surface of the 3D model are grouped to a surface region which is textured with one common image in a first step. In a second step the boundaries between neighbouring surface regions are filtered with a local texture filter. The texture for surface regions which are invisible in any image is synthesized from the texture of neighbouring surface regions in a last step. Applying this method leads to less distorted textured models.

Results with real image sequences have confirmed the suitability of the proposed algorithm for convex objects with complex and detailed surfaces. Though the quality of resulting concave models mainly depends on the quality of the integrated depth information. Due to the newly developed texture estimation algorithm the virtual models receive a natural look. Small details of the object which could not be modelled in the 3D geometry are well represented in the texture.

In future work the system will be extended by the facility to model outdoor objects by steering a camera around the object.

8 Acknowledgements

This work was supported by the RACE II project R2052 ”MONA LISA”. Partners in this project are Thomson Broadband Systems, BBC, Daimler Benz, DVS Digitale Videosysteme GmbH, Queen Mary & Westfield College, Siemens AG, University of Balearic Islands, and VAP Video Art Production GmbH.

9 References

[1] A. Blake, A. Zisserman. ”*Visual Reconstruction*”, MIT Press, 1987

[2] H. Busch, ”Ein Verfahren zur Oberflächenmodellierung dreidimensionaler Objekte aus Kamerabildfolgen”, *Ph.D. thesis*, *Universität Hannover*, 1991.

[3] C. H. Chien and J. K. Aggarwal, ”Identification of 3D objects from multiple silhouettes using quadtrees/octrees ”, *Comp. Vision, Graphics, and Image Processing* 36, 1986, pp. 256–273.

[4] L. Falkenhagen, ”Depth Estimation from Stereoscopic Image Pairs assuming Piecewise Continuos Surfaces”, *European Workshop on Combined real and synthetic image processing for broadcast and video productions*, Hamburg, Germany, November 1994.

[5] J.D. Foley et al., ”*Computer Graphics: Principles and Practice*”, Addison Wesley, 1992.

[6] Koch, R., ”Automatic Reconstruction of Buildings from Stereoscopic Image Sequences”, *Eurographics ’93*, Barcelona,Spain, 1993.

[7] W. N. Martin and J. K. Aggarwal , ‘‘Volumetric Descriptions of Objects from Multiple Views” ,*Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI–5, 1983, pp. 150–159.

[8] D. Meagher, ”Geometric modelling using octree encoding”, *Computer Vision, Graphics, and Image Processing*, Vol 19, 1982, pp. 129–147.

[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.

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

[11] R. Szeliski, ”Rapid Octree Construction from Image Sequences”, *CVGIP: Image Understanding*, Vol. 58, No 1, July, pp. 23–32, 1993.

[12] D. Terzopoulos, ”The Computation of Visible–Surface Representations”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, July 1988.

[13] R. Y. Tsai, ”A Versatile Camera Calibration Technique for High–Accuracy 3D Machine Vision Metrology Using Off–the–Shelf TV Cameras and Lenses”, *Journal of Robotics and Automation*, Vol. RA–3. No.4, August 1987, pp. 323–344.

[14] Y.Yakimovsky, R. Cunningham, ”A System for Extracting Three–Dimensional Measurements from a Stereo Pair of TV Cameras”, *Computer Vision, Graphics, and Image Processing*, Vol. 7, pp. 195 –210, 1978.