VOLUMETRIC MODELING USING SHAPE FROM SILHOUETTE

Yasemin Kuzu and Volker Rodehorst

Photogrammetry and Cartography, Technical University of Berlin, Germany

In this paper we describe an efficient image-based approach to compute volume models from silhouette images in a low cost measuring environment. First, the contours of a real object must be extracted from a sequence of images.
In a second step, a bounding pyramid is constructed using the projection center and the silhouette. The intersection of all bounding pyramids defines an approximate geometric representation for the object called visual hull.
Silhouette based constructions have the principal inability to detect the object concavities. We present some new results to minimize the modeling error on critical areas using a block matching approach.

1. INTRODUCTION

Computer Vision concerns with the automatic extraction of various information from imagery. One of its purposes is the creation of computer models of 3-D objects from digital images. Furthermore, obtaining precise measurements from photographs has a long history in photogrammetry. In this paper, we describe an efficient image-based approach to compute volume models from silhouette image data.

There is a growing need for an automated generation of 3-D computer models in many applications. They can be used as the first input stage for designer, computer graphic artists and virtual reality experts. The 3-D object acquisition can be performed by range sensors such as laser scanners. Using image sequences of a monoscopic camera is a low cost alternative and therefore attractive for all these applications.

Modeling the shape of real world objects from a series of images has been extensively studied in the last years. One of the well-known approaches for 3-D modeling is shape from silhouette, which recovers the shape of the objects from their contours. This approach is popular due to its fast computation and robustness.

The first work on the construction of 3-D models from multiple views has been done by *Martin and Aggrawal* [1983]. *Chien and Aggrawal* [1986] used orthographic projection for the construction of volume models. *Potmesil* [1987] and *Srivastava and Ahuja* [1990] used arbitrary views and perspective projection. Numerous researchers have dealt with the shape from silhouette technique to convert visible contours into a visual hull, i.e. [*Szeliski*, 1991], [*Niem*, 1994], [*Kutulakos and Seitz*, 1998] and [*Vedula et al.*, 1998].

However, the concavities on an object can not be recovered from large number of views since the viewing region doesn't completely surround the object. In this paper we present some new results to detect critical areas and support the shape from silhouette with an area based stereo technique to solve this problem. Applying a block matching algorithm, the correlation of the surface texture is used to get inside the visual hull.

2. SYSTEM CONFIGURATION

2.1 Experimental Setup

The experimental system configuration use a single stationary video CCD camera and a turntable as a controlled motion platform on which the object is placed (see Fig.1).

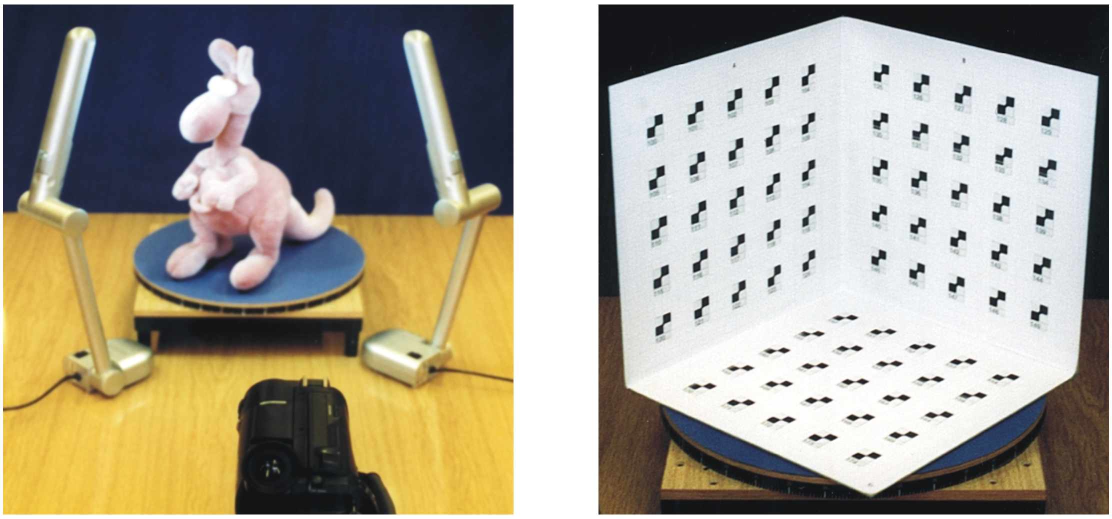


Figure 1: Experimental setup using a turntable (left) and a calibration object (right) in front of a monoscopic camera.

2.2 Camera Calibration

Before starting the image acquisition, the system must be calibrated. The interior and exterior parameters of the camera must be known, i.e. the position and orientation of the camera and the focal length. We perform a system calibration by using a special calibration object (cp. Fig.1). It provides a good coverage of the object with three square faces containing 25 control points on each side. The geometric description of the configuration using central projection without the a lens distortion model is shown in Fig. 2:

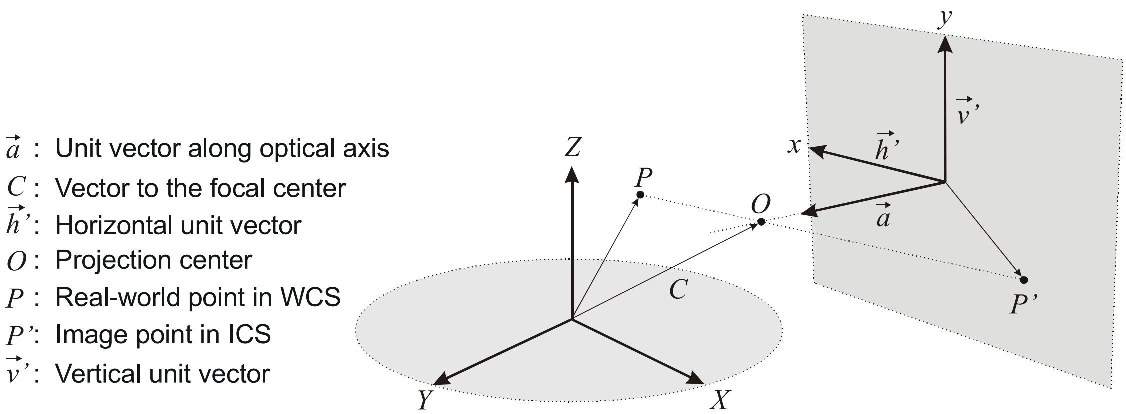


Figure 2: Mathematical camera model for central projection.

3. AUTOMATIC IMAGE SEGMENTATION

3.1 Background Estimation

In our experiment, the turntable is rotated and the video images are captured and pre-processed. First, the contour of the real object must be extracted from the input images. Therefore, a monochromatic background was used to distinguish the object from the environment. For a small frame (see dotted line in Fig.3) the background color is estimated using a histogram of the hue values.

Figure 3: Background estimation using an IHS color space histogram (left) and the resulting silhouette extraction (right).

3.2 Image Segmentation

For every color pixel (R,G,B) a transformation into the IHS color space is computed:

$$I = \frac{R+G+B}{3} \qquad S = 1 - \frac{\min(R,G,B)}{I}, \quad I > 0.05$$

$$H = \arccos\left(\frac{(R-G)+(R-B)}{\sqrt{(R-G)^2 + (R-B)\cdot(G-B)}} \right), \quad B < G$$

The automatic segmentation algorithm compares all the hue data with the estimated background information using a simple threshold and a minimum saturation value. Finally, a morphological operation ('opening') eliminates isolated pixel.

4. SHAPE MODELLING USING VOXEL CARVING

4.1 The Visual Hull

When the camera geometry is known, a bounding pyramid can be constructed for every image. This is performed by the lines of sight from the camera focal point through all contour points of the object silhouette. The shape is computed volumetrically by carving away all voxels outside the projected silhouette cone (see Fig.4). The intersection of all silhouette cones from multiple images defines an estimate geometry of the object called visual hull [*Laurentini*, 1995, *Matusik et al*, 2000].

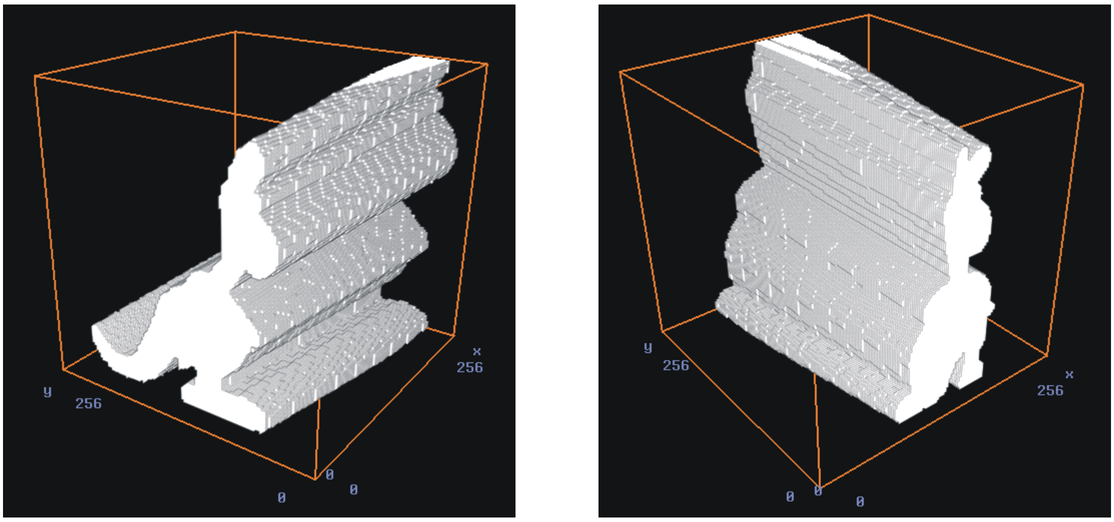


Figure 4: Voting-based carving of a voxel cube using various silhouettes under central projection.

4.2 Voting-based Voxel Carving

The carving algorithm use the position for each voxel in real-world coordinates P and perform a projection into image coordinates P' using the following equations:

$$P'_x = \frac{(P-C)\cdot\vec{h}}{(P-C)\cdot\vec{a}} \qquad P'_y = \frac{(P-C)\cdot\vec{v}}{(P-C)\cdot\vec{a}} \qquad \vec{h} = f\cdot\vec{h}' + \vec{a}\cdot x_h \qquad \vec{v} = f\cdot\vec{v}' + \vec{a}\cdot y_h$$

If the image coordinate defines a background pixel, the voxel is marked to be deleted (voting). After the processing of all voxels, the cube is rotated and the algorithm continue with the next image. When the greater numbers of views are used, this technique progressively refines the object model. Finally, the voxels are purged using a threshold for the number of votes.

The Fig.5 shows an experimental result for the visual hull with a real toy kangaroo using the presented voxel carving algorithm.

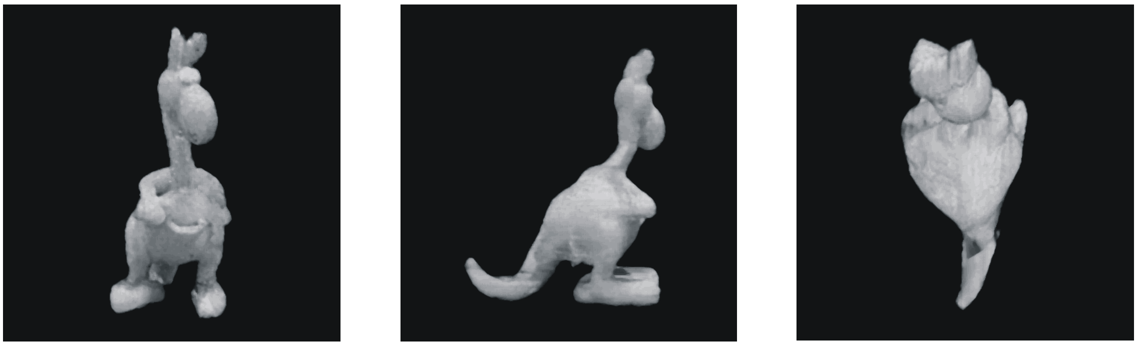


Figure 5: Voxel model of the 3-D shape reconstruction using 60 images and 256^3 voxels with texture mapping.

5. HIERARCHICAL PROCESSING USING OCTREES

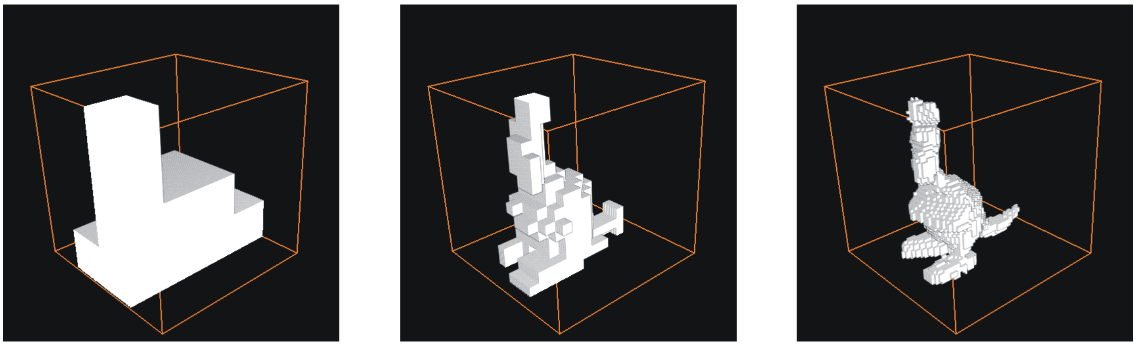


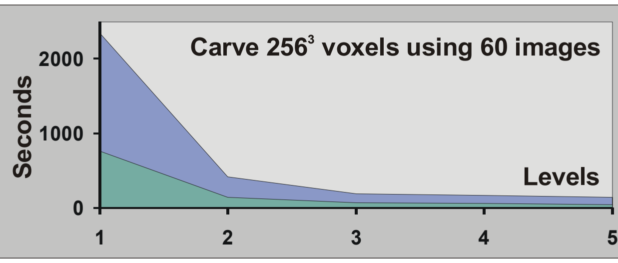
Figure 6: Iterative generation of an octree model using 4^3, 16^3, and 64^3 voxels.

We enhanced the voted-based voxel carving by using a simple octree structure [*Srivastava and Ahuja*, 1990, *Potmesil*, 1987]. The voxels at one level can be derived from the results of the preceding level by applying a modified algorithm. Some examples of the iterative generation are shown in Fig. 6. The restriction to areas, which contain relevant object information, accelerate the computation time significantly (see. Tab. 1).

Table 1: Speed optimization using hierarchical processing.

Level	Intel PIII	SGI R12k
1	39:00	12:40
2	7:01	2:21
3	3:05	1:03
5	2:26	0:51
7	2:16	0:48

Computation time (in min.)



6. STEREO ANALYSIS USING BLOCK-MATCHING

6.1 Polychromatic Block Matching

The modeling error of the visual hull can be minimized using a block matching approach [*Koschan*, 1993]. If the surface contains enough texture, the missing object concavities can be detected. The main idea of Block Matching is a similarity check between two equal sized blocks in two images (area-based stereo). The mean square error MSE between the pixel values inside the blocks defines a measure for the similarity. The block is shifted pixel by pixel inside a search area, that is defined by the epipolar geometry.

$$D = \min_{|\Delta| \leq d_{MAX}} \{MSE_{COLOR}(x, y, \Delta)\}$$

The block disparities D are median filtered to avoid outliers and a dense depth map is generated using a pixel selection technique.

Figure 7: Computed stereo depth maps of the flower (left, middle) and the combination of three neighbored views (right).

6.2 Hierarchical Matching Using Image Pyramids

The enhanced algorithm [*Koschan and Rodehorst*, 1997] is more robust and shows better results. The hierarchical approach can be implemented very efficiently in parallel to achieve high speed execution. Furthermore, the combination of three views improves the quality of the matching results (see Fig.7). The correspondence analysis of a reference image with the neighbored images eliminates most of the artifacts.

6.3 Integration of Volume Data and Depth Maps

The combination of the voxel data and the depth maps is realized in the voted-based carving stage. Each voxel get additional votes according to the information provided by the depth maps. First results (see Fig.8) show the importance of precise calibration data and additional constraints are necessary to reduce the number of outliers.

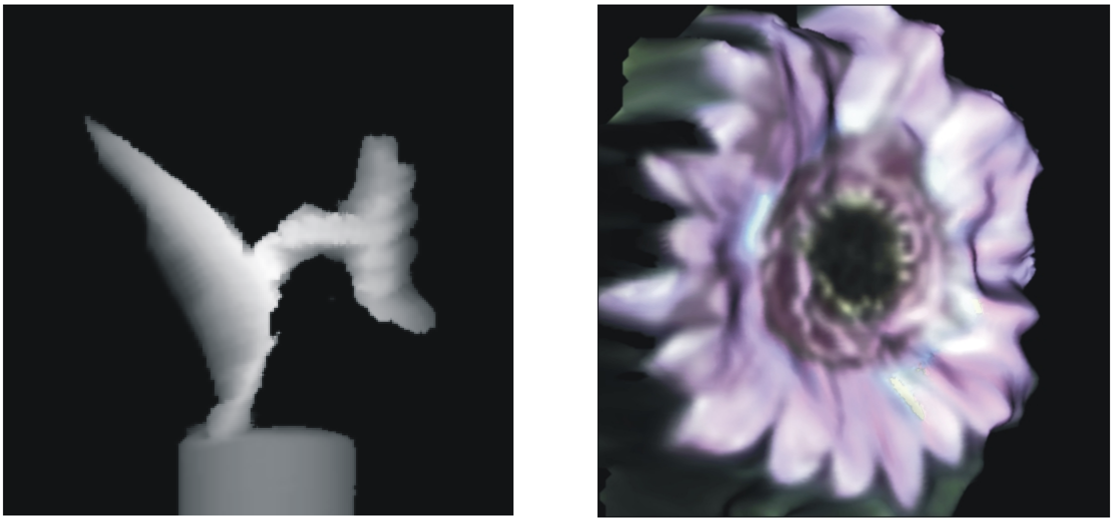


Figure 8: Voxel model of the flower (left) and the enhanced concave head using block matching (right).

7. SUMMARY AND FUTURE WORK

We presented an approach to construct 3-D models of real-world objects using silhouette data in a controlled environment. The system requirements are simple therefore attractive for many applications.

The silhouette extraction method using histogram information in the IHS color space works robust. The hierarchical processing reduces the computation time significantly. The combination with the block matching technique to handle concavities still suffers from a lack in accuracy. Additional tests and investigations are necessary, but further improvements are in progress and will be presented soon. The future work will concentrate on auto calibration, error estimation and color texture mapping.

Acknowledgements. The authors would like to thank J. Moré and A. Wiedemann for their helpful assistance to the camera calibration task.

REFERENCES

C. H.Chien and J. K. Aggarwal, “Identification of 3D Objects from Multiple Silhouettes Using Quadtrees / Octrees“, Computer Vision Graphics And Image Processing 36, pp. 256-273 (1986).

A.W. Fitzgibbon, G. Cross and A. Zissermann, “Automatic 3D Model Construction for Turntable Sequences”, Proceedings of the European Workshop on 3D Structure from Multiple Images of Large-scale Environments (SMILE '98), LNCS 1506, pp. 155-170 (1998).

A. Koschan, “Dense Stereo Correspondence Using Polychromatic Block Matching”, Proc. of the 5th Int. Conf. on Computer Analysis of Images and Patterns (CAIP’93), D. Chetverikov, W. Kropatsch (eds.), Budapest, Hungary, pp. 538-542 (1993).

A. Koschan and V. Rodehorst, “Dense Depth Maps by Active Color Illumination and Image Pyramids”, Advances in Computer Vision, F. Solina, W.G. Kropatsch, R. Klette, R. Bajcsy (eds.), Springer, pp. 137-148, (1997).

N. Kutulakos and M. Seitz, “A Theory of Shape by Space Carving”, University of Rochester CS Technical Report 692 (1998).

A. Laurentini, “How far 3D Shapes Can Be Understood from 2D Silhouettes”, IEEE Transactions on Pattern Analysis and Machine Intelligence”, Vol.17, No.2 (1995).

W. N. Martin and J. K. Aggarwal, “Volumetric Descriptions of Objects from Multiple Views”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-5, No.2 (1983).

W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan "Image-Based Visual Hulls", SIGGRAPH 2000 , Computer Graphics Proceedings, Annual Conference Series, pp. 369-374 (2000).

W. Niem, "Robust and Fast Modeling of 3D Natural Objects from Multiple Views", Proceedings "Image and Video Processing II", Vol. 2182, pp. 388-397 (1994).

M. Potmesil, “Generating Octree Models of 3D Objects from Their Silhouettes in a Sequence of Images”, Computer Vision Graphics And Image Processing 40, pp. 1-29 (1987).

S. K. Srivastava and N. Ahuja, “Octree Generation from Object Silhouettes in Perspective Views”, Computer Vision, Graphics And Image Processing 49, pp. 68-84 (1990).

R. Szeliski, ”From images to models (and beyond): a personal retrospective”, Vision Interface '97, Kelowna, British Columbia, Canadian Image Processing and Pattern Recognition Society, pp. 126-137 (1997).

R. Szeliski, “Shape from rotation”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91), pp. 625-630 (1991).

S. Vedula, P. Rander, H. Saito and T. Kanade, “Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences”, Proc. 4th Conference on Virtual Systems and Multimedia (VSMM98), pp. 326-332 (1998).