Volume Estimation for Objects with Concavities *

Srdan Tosovic and Robert Sablatnig

Pattern Recognition and Image Processing Group,
Institute for Computer Aided Automation,
Vienna University of Technology,
Favoritenstraße 9/183/2, A-1040 Vienna, Austria
{tos,sab}prip.tuwien.ac.at, http://www.prip.tuwien.ac.at

Abstract. An algorithm for the automatic construction of a 3D model of archaeological vessels is presented. In archeology the determination of the exact volume of arbitrary vessels is of importance since this provides information about the manufacturer and the usage of the vessel. To acquire the shape of objects with handles in 3d is complicated, since occlusions of the object's surface are introduced by the handle and can only be resolved by taking multiple views. Therefore, the 3d reconstruction is based on a sequence of images of the object taken from different viewpoints. The object's silhouette is the only feature which is extracted from an input image. Images are acquired by rotating the object on a turntable in front of a stationary camera. The algorithm uses an octree representation of the model, and builds this model incrementally, by performing limited processing of all input images for each level of the octree. Results of the algorithm developed are presented for both synthetic and real input images.

1 Introduction

The *Shape from Silhouette* method presented in this paper was performed within the *Computer Aided Classification of Ceramics* [MS96,KS99] project, which aims to provide an objective and automated method for classification and reconstruction of archaeological pottery. The final goal is to provide a tool which helps archaeologists in their classification process.

Pottery was made in a very wide range of forms and shapes. The purpose of classification is to get a systematic view of the material found, to recognize types, and to add labels for additional information as a measure of quantity [OTV93]. In this context, decoration of pottery is of great interest. Decoration is difficult to illustrate since it is a perspective projection of an originally spherical surface. In order to be able to unwrap the surface it is necessary to have a 3d representation of the original surface. Furthermore, the exact volume of the vessel is of great

* This work was partly supported by the Austrian Science Foundation (FWF) under grant P13385-INF.

interest to archaeologists too, since the volume estimation also allows a more precise classification [OTV93].

Since pottery is manufactured on a turntable we use a turntable based method for the 3d-reconstruction of the original. To acquire images from multiple views we put the archaeological vessel on a turntable which rotates in front of a stationary camera. Shape from Silhouette is a method of automatic construction of a 3D model of an object based on a sequence of images of the object taken from multiple views. In these views the object's silhouette represents the only interesting feature [Sze93,Pot87]. The object's silhouette in each input image corresponds to a conic volume in the object's real-world space (Figure 1). A 3D model of the object can be built by intersecting the conic volumes from all views, which is also called *Space Carving* [VBSK00].

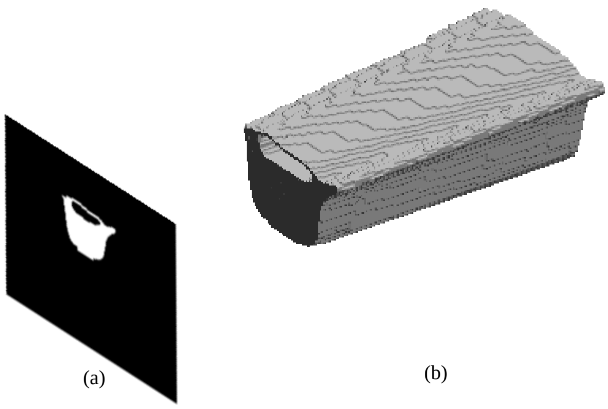


Fig. 1. Image silhouette (a) and the corresponding conic volume (b)

Shape from Silhouette is a computationally simple algorithm (it employs only basic matrix operations for all transformations) and it requires only a camera as equipment, so it can be used to obtain a quick initial model of an object which can then be refined by other methods like active triangulation [FL98]. It can be applied to objects of arbitrary shapes, including objects with certain concavities (like a handle of a cup), as long as the concavities are visible from at least one input view. It can also be used to estimate the volume of an object.

There have been many works on construction of 3D models of objects from multiple views. Baker [Bak77] used silhouettes of an object rotating on a turntable to construct a wire-frame model of the object. Martin and Aggarwal [MA83] constructed volume segment models from orthographic projection of silhouettes. Chien and Aggarwal [CA83] constructed an object's octree model from its three orthographic projections. Veenstra and Ahuja [VA86] extended this approach to thirteen standard orthographic views. Potmesil [Pot87] created octree models using arbitrary views and perspective projection. For each of the views he constructs an octree representing the corresponding conic volume (Figure 1) and then intersects all octrees. In contrast to this, Szeliski [Sze93] first creates a low resolution octree model quickly and then refines this model iteratively, by intersecting each new silhouette with the already existing model. The last two approaches project an octree node into the image plane to perform the intersection between the octree node and the object's silhouette. Srivastava and Ahuja [SA90] in contrast, perform the intersections in 3D-space. The works of Szeliski [Sze93] and Niem [Nie99] were used as a basis for the approach presented in this paper.

2 Acquisition System

The acquisition system [KT00] consists of the following devices:

- a monochrome CCD-camera with a focal length of 16 mm and a resolution of 768x576 pixels
- a turntable with a diameter of 50 cm, whose desired position can be specified with an accuracy of 0.05 degrees

The distance between the camera and the turntable is approx. 120 cm and is estimated by the calibration algorithm. Figure 2 depicts the geometrical setup of the camera and the turntable.

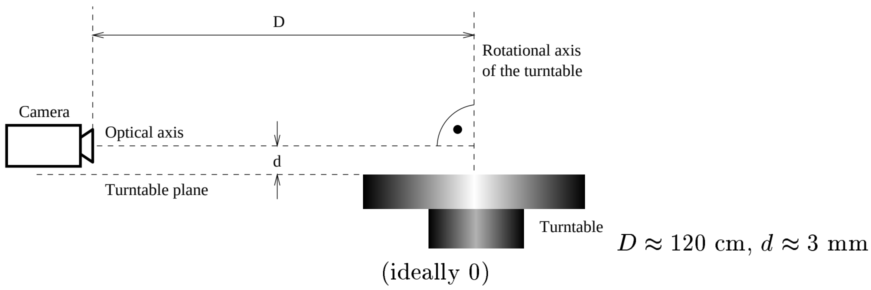


Fig. 2. Geometrical setup of the camera and the turntable

An important issue is the illumination of the object observed, which should be clearly distinguishable from the background, independent from the object's shape or the type of its surface. For that reason back-lighting [HS91] is used. A large (approx. 50x40 cm) rectangular lamp is put behind the turntable (as seen from the camera). In addition, a white piece of paper, larger than the lamp, is put right in front of the lamp, in order to make the light more diffuse. The whole system is protected against the ambient light by a thick black curtain. Prior to any acquisition, the system is calibrated in order to determine the inner and outer orientation of the camera and the rotational axis of the turntable. The calibration method used was exclusively developed for the Shape from Silhouette algorithm presented and is described in detail in [Tos99] and [KT00].

3 Model Representation

Many different model representations are used in computer vision and computer graphics. Here we will mention only the one used in our case An octree [CH88] is a tree-formed data structure used to represent 3-dimensional objects. Each node of an octree represents a cube subset of a 3-dimensional volume. A node of an octree which represents a 3D object is said to be:

- *black*, if the corresponding cube lies completely within the object
- *white*, if the corresponding cube lies completely within the background, i.e., has no intersection with the object
- *gray*, if the corresponding cube is a boundary cube, i.e., belongs partly to the object and partly to the background. In this case the node is divided into 8 child nodes (octants) representing 8 equally sized sub-cubes of the original cube

All leaf nodes are either black or white and all intermediate nodes are gray. An example of a simple 3D object and the corresponding octree is shown in Figure 3.

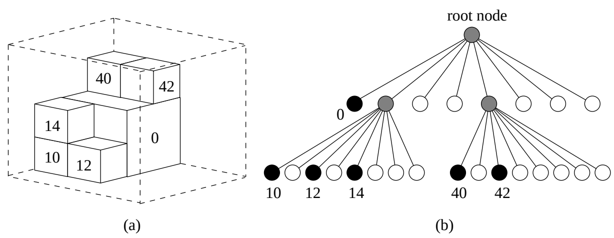


Fig. 3. A simple object (a) and the corresponding octree (b)

The octree representation has several advantages [CH88]: for a typical solid object it is an efficient representation, because of a large degree of coherence between neighboring volume elements (voxels), which means that a large piece of an object can be represented by a single octree node. Another advantage is the ease of performing geometrical transformations on a node, because they only need to be performed on the node's vertices. The disadvantage of octree models is that they digitize the space by representing it through cubes whose resolution depends on the maximal octree depth and therefore cannot have smooth surfaces.

4 Shape from Silhouette

The algorithm builds up a 3D model of an object in the following way: first, all input images are transformed into binary images where a "black" pixel belongs to the object observed and a "white" one to the background[1] (Figure 4a). Then, the initial octree is created with a single root node (Figure 4b) representing the whole object space, which will be "carved out" corresponding to the shape of the object observed.

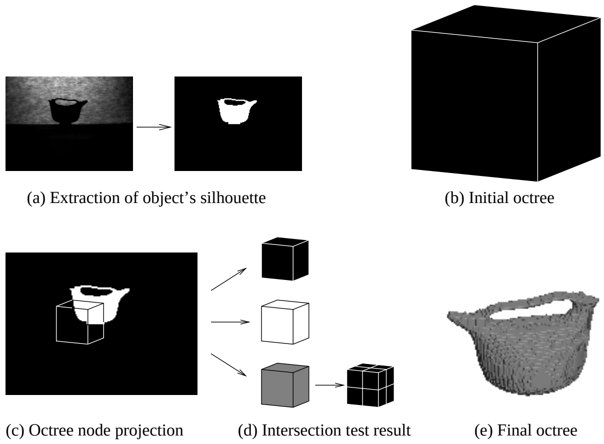


Fig. 4. Algorithm overview

Next, the octree is processed in a level-by-level manner: starting from level 0 (with root node as the only node), all octree nodes of the current level marked as "black", i.e., belonging to the object, are projected into the first input image (Figure 4c) and tested for intersection with the object's image silhouette. Depending on the result of the intersection test, a node can remain to be "black", it can be marked as "white" (belonging to the background) or in case it belongs partly to the object and partly to the background, it is marked as "grey" and divided into 8 black child nodes of the next higher level (Figure 4d). The remainder of the black nodes of the current level is then projected into the next input image where the procedure of intersection testing with the object's silhouette is repeated. Once all input images have been processed for the current octree level, the current level is incremented by one and the whole procedure (starting from the projection of the black nodes of the current level into the first input image) is repeated, until the maximal octree level has been reached. The remaining octree after the processing of the last level is the final 3D model of the object (Figure 4e).

---

[1] in the implementation, *black* means background and *white* means object, but it is more intuitive to describe an object pixel as "black" and a background pixel as "white"

5 Results

We performed tests with two synthetic 3D objects: a sphere and a cone. Both objects are rotational-symmetric, so if we assume that they rotate around their axis of symmetry, we need to create only one synthetic input image for each object, because the projection of the object into the image plane will always look the same, independent of the viewing angle. We assume we have such a virtual acquisition system where:

- a millimeter in the x–y plane of the object coordinate system corresponds to exactly one pixel in the image plane
- x and y axes of the object coordinate system are parallel to the x and y axes of the image coordinate system

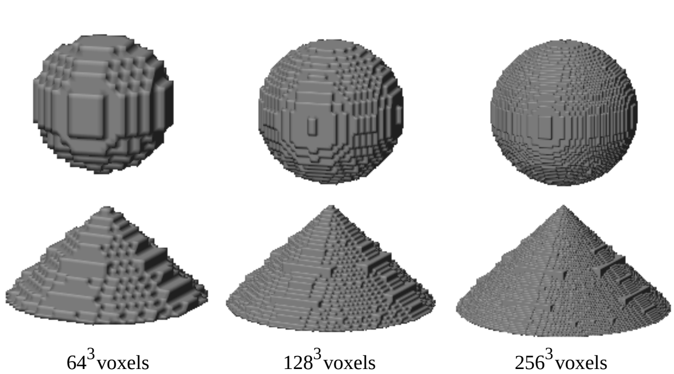


Fig. 5. Constructed models of synthetic objects in different voxel resolutions

Under these assumptions we can construct perfect synthetic input images for the sphere and the cone, i.e., the transformation between the virtual image and object coordinate systems will be described by a synthetic transformation matrix T without any errors. With the results of the tests of these synthetic images we can estimate the accuracy of the Shape from Silhouette algorithm implemented, without any influence of camera calibration errors, which is unavoidable for a real acquisition system. In the first sequence of tests we construct a 3D model of an object based on 36 different views, with an angle of 10° between two consequent views, in 64^3, 128^3 and 256^3 voxel resolutions. Figure 5 shows the reconstructed 3D models of our synthetic objects for each of these resolutions and Table 1 compares their analytic and the computed dimensions and volume (the voxel size is given in mm).

For tests with real objects we used a metal cuboid and two ceramic pots (Figure 6). The acquisition system was calibrated resulting in a calculated distance of 118.0 cm between the camera and the rotation axis of the turntable (D in Figure 2).

object	dimensions (mm)	voxel size	volume (mm^3)
sphere	120x120x120	—	904 779
64^3 v.	112x112x112	4	733 184 (-18.96%)
128^3 v.	120x116x120	2	823 296 (-9.01%)
256^3 v.	120x120x120	1	883 712 (-2.33%)
cone	250x125x250	—	2 045 308
64^3 v.	224x112x224	4	1 802 240 (-11.88%)
128^3 v.	240x120x240	2	1 943 040 (-5.00%)
256^3 v.	248x124x248	1	2 037 056 (-0.40%)

Table 1. Comparison of analytic and calculated dimensions

(a) (b) (c)

Fig. 6. Real objects: a metal cuboid (a) and two ceramic pots (b) and (c)

In a test sequence we built three octree models for each object in Figure 6, by using 4, 12 and 36 input views, with a uniform angle between two consequent views, with an octree resolution of 256^3 voxels. Figure 7 shows the 3D models constructed and Table 2 gives the octree statistics, i.e., its maximal width, height and depth with the maximal width, height and depth of the reconstructed 3D model, the voxel size, the number of nodes in the octree and the CPU time on a Pentium II 450 MHz with 256 MB of RAM. .

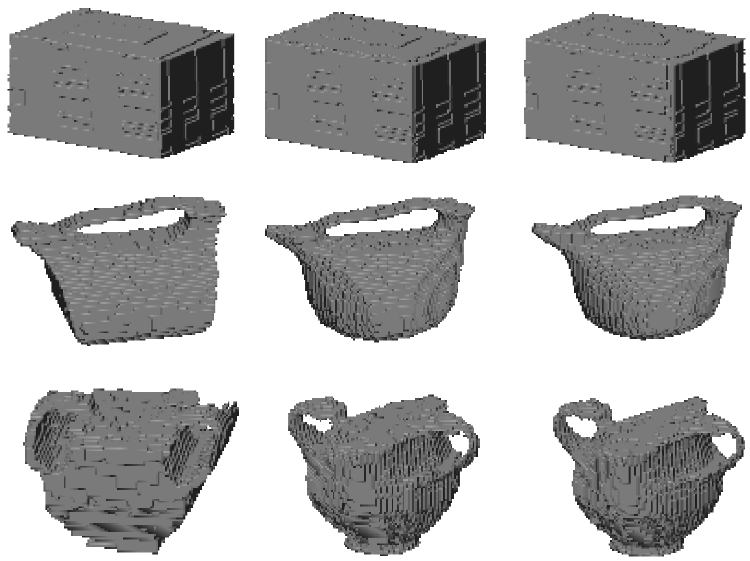
Finally, Figure 8 shows the reconstructed 3D models of the two pots from three sides. For these models an octree of a resolution of 256^3 voxels was built, based on input images from 36 views.

The results with both synthetic and real input data show that there is a certain minimal octree resolution required to obtain an accurate model of an object, especially for highly detailed objects, like the two pots used for tests with real images. To increase the octree resolution to 512^3 would not improve the results of our tests, because the projection of a single voxel would be less than half the pixel's size. Concerning the number of input views used for obtaining a model of an object, it turned out that beginning from 12 views, the constructed model does not change significantly. In our tests the octrees built from 12 views were almost the same as the ones built from 36 views, except that they took much less time to construct.

The results with synthetic data, where we had a perfect transformation matrix, showed that the error in the dimensions of the model lies within or is slightly higher than the error introduced through the minimal voxel size. The error with

object	dimensions (mm)	voxel size	# of nodes	CPU time
cuboid	100.2x60.1x70.3	—	—	—
4 views	109.0x62.1x77.3	0.59	84 065	9.76 s
12 views	109.0x62.1x77.3	0.59	161 625	15.15 s
36 views	109.0x62.1x77.3	0.59	310 649	36.10 s
pot 1	141.2x93.7x84.8	—	—	—
4 views	148.4x93.8x89.1	0.78	275 177	12.30 s
12 views	145.3x93.8x89.1	0.78	486 585	23.47 s
36 views	145.3x93.8x85.9	0.78	857 753	67.28 s
pot 2	114.2x87.4x114.6	—	—	—
4 views	118.8x89.1x115.6	0.78	257 897	12.57 s
12 views	117.2x89.1x114.1	0.78	424 201	23.32 s
36 views	117.2x89.1x114.1	0.78	726 689	65.10 s

Table 2. Octree statistics for real data with varying numbers of views



4 views 12 views 36 views

Fig. 7. Constructed models of real objects with different number of views

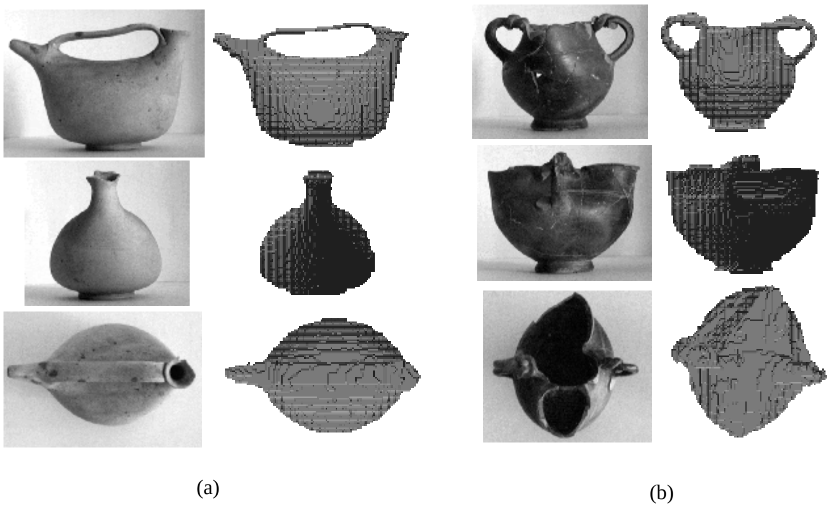


Fig. 8. Reconstructed 3D models of the pot 1 (a) and pot 2 (b)

real data depends additionally on the accuracy of the calibration algorithm. The results also showed that the algorithm works much better with oval objects, i.e., with objects that do not have completely flat surfaces or sharp edges.

6 Conclusions and Future Work

In this report an implementation of Shape from Silhouette method was presented, which creates a 3D model of an object from images of the object taken from different viewpoints. It showed to be a simple and fast algorithm, which is able to reconstruct models of arbitrarily shaped objects, as long as they do not have too many hidden concavities, i.e., concavities not visible in any of the input images. The algorithm is simple, because it employs only simple matrix operations for all the transformations and it is fast, because even for highly detailed objects, a high resolution octree (256^3 voxels) and a high number of input views (36), the computational time hardly exceeded 1 minute on a Pentium II. Already for a smaller number of views (12) the constructed models were very similar to the ones constructed from 36 views and they took less than 25 seconds of computational time.

For archaeological applications, the object surface has to be smoothed in order to be applicable to ceramic documentation, for classification, however, the accuracy of the method presented is sufficient since the projection of the decoration can be calculated and the volume estimation is much more precise than the estimated volume performed by archaeologists.

A drawback of the Shape from Silhouette algorithm presented consists in using "too much" knowledge about the calibration. For the projection of the nodes into the image plane it uses the fact that the image plane and the x–y plane of the object coordinate system are completely parallel, making the algorithm unusable for any other setup of camera and turntable, which is currently under development, where the camera, for example, looks at the turntable from the angle of 45°. Furthermore, the method of acquiring binary images with object's silhouette from a real-world input image is improved by using edges rather than binary images.

References

[Bak77] H. Baker. Three-dimensional modelling. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pages 649–655, 1977.

[CA83] C. H. Chien and J. K. Aggarwal. Volume/surface octrees for the representation of three-dimensional objects. *Computer Vision, Graphics, and Image Processing*, 36:100–113, 1983.

[CH88] H. H. Chen and T. S. Huang. A survey of construction and manipulation of octrees. *Computer Vision, Graphics, and Image Processing*, 43:409–431, 1988.

[FL98] C.S. Fuh and H.B. Liu. Projection for pattern recognition. *IVC*, 16(9-10):677–687, July 1998.

[HS91] R. M. Haralick and L. G. Shapiro. Glossary of computer vision terms. *Pattern Recognition*, 24(1):69–93, 1991.

[KS99] M. Kampel and R. Sablatnig. On 3d Modelling of Archaeological Sherds. In *Proceedings of International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, Santorini, Greece*, pages 95–98, 1999.

[KT00] M. Kampel and S. Tosovic. Turntable calibration for automatic 3D-reconstruction. In *Applications of 3D-Imaging and Graph-based Modelling, Proceedings of the 24th Workshop of the Austrian Association for Pattern Recognition (ÖAGM)*, pages 25–31, 2000.

[MA83] W. N. Martin and J. K. Aggarwal. Volumetric description of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2):150–158, 1983.

[MS96] C. Menard and R. Sablatnig. Computer based Acqusition of Archaeological Finds: The First Step towards Automatic Classification. In Hans Kamermans and Kelly Fennema, editors, *Interfacing the Past, Computer Applications and Quantitative Methods in Archaeology*, number 28, pages 413–424, Leiden, March 1996. Analecta Praehistorica Leidensia.

[Nal93] V. S. Nalwa. *A Guided Tour Of Computer Vision*. Addison-Wesley, 1993.

[Nie99] W. Niem. Automatic reconstruction of 3d objects using a mobile camerahle. *IVC*, 17(2):125–134, February 1999.

[OB79] J. O'Rourke and N. Badler. Decomposition of three-dimensional objects into spheres. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(3):295–305, 1979.

[OTV93] C. Orton, P. Tyers, and A. Vince. *Pottery in Archaeology*, 1993.

[Pot87] M. Potmesil. Generating octree models of 3D objects from their silhouettes in a sequence of images. *Computer Vision, Graphics, and Image Processing*, 40:1–29, 1987.

[SA90] S. K. Srivastava and N. Ahuja. Octree generation from object silhouettes in perspective views. *Computer Vision, Graphics, and Image Processing*, 49:68–84, 1990.

[Shi87] Y. Shirai. *Three-Dimensional Computer Vision*. Springer-Verlag, 1987.

[Sze93] R. Szeliski. Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58(1):23–32, July 1993.

[Tos99] S. Tosovic. Lineare Hough-Transformation und Drehtellerkalibrierung. Technical Report PRIP-TR-59, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group, Vienna University of Technology, Austria, 1999.

[VA86] J. Veenstra and N. Ahuja. Efficient octree generation from silhouettes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 537–542, 1986.

[VBSK00] S. Vedula, S. Baker, S. Seitz, and T. Kanade. Shape and motion carving in 6d. In *CVPR00*, volume 2, pages 592–598, 2000.