

# Fusion of surface and volume data <sup>1)</sup>

Martin Kampel, Robert Sablatnig and Srdan Tosovic

Vienna University of Technology,

Institute of Computer Aided Automation,

Pattern Recognition and Image Processing Group

Favoritenstr. 9, 183-2, A-1040 Vienna, Austria

{kempel,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 using two different 3d algorithms 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 3d shape of objects with handles 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 with two different algorithms; shape from silhouette and shape from structured light. The output of both algorithms are then used to construct a single 3d model. Results of the algorithm developed are presented for both synthetic and real input images.*

## 1 Introduction

The combination of the *Shape from Silhouette* (SfS) method with the *Shape from structured Light* (SfL) method presented in this paper was performed within the *Computer Aided Classification of Ceramics* [6] project, which aims to provide an objective and automated method for classification and reconstruction of archaeological pottery. The volume of the vessel is of interest to archaeologists, since the volume estimation allows a more precise classification.

SfS 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 which the object's silhouette represents the only interesting feature of the image [2, 12]. The object's silhouette in each input image corresponds to a conic volume in the object real-world space. A 3d model of the object can be built by intersecting the conic volumes from all views, which is also called *Space Carving* [7]. The method can be applied on objects of arbitrary shapes, including objects with certain concavities, as long as the concavities are visible from at least one input view [14]. This condition is hard to hold since most of the ar-

---

<sup>1)</sup>This work was partly supported by the Austrian Science Foundation (FWF) under grant P13385-INF, the EU under grant IST-1999-20273, and the Austrian Federal Ministry of Education, Science and Culture.

archaeological vessels do have concavities that have to be modeled. Therefore, a second, active shape determination method has to be used to discover all concavities. The second acquisition method used is SfL, based on active triangulation [3, 4].

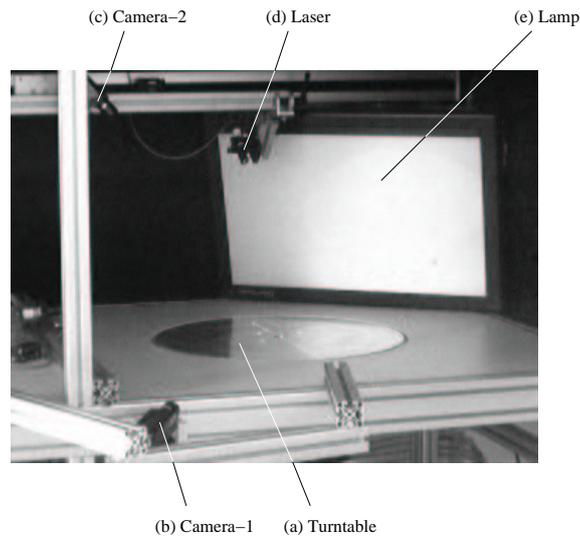
There have been many works on construction of 3d models of objects from multiple views. Baker [1] used silhouettes of an object rotating on a turntable to construct a wire-frame model of the object. Martin and Aggarwal [10] constructed volume segment models from orthographic projection of silhouettes. Potmesil [12] created octree models using arbitrary views and perspective projection. In contrast to this, Szeliski [13] first created a low resolution octree model quickly and then refined this model iteratively, by intersecting each new silhouette with the already existing model. The work of Szeliski [13] and Niem [11] was used as a basis for the SfS approach presented in this paper. For the active triangulation method we use an approach by Liska developed for a next view planing strategy using structured light [9].

## 2 Acquisition System

The acquisition system consists of the following devices:

- a turntable (Figure 1a) with a diameter of 50 *cm*, and a positional accuracy of  $0.05^\circ$ .
- two monochrome CCD-cameras with a focal length of 16 *mm* and a resolution of 768x576 pixels. One camera (*Camera-1* in Figure 1) is used for acquiring the images of the object's silhouettes and the other (*Camera-2* in Figure 1) for the acquisition of the images of the laser light projected onto the object.
- a red laser (Figure 1d) used to project a light plane onto the object. The laser is equipped with a prism in order to span a plane out of the laser beam.
- a lamp (Figure 1e) used to back-light [5] the scene for the acquisition of the silhouette of the object.

Both cameras are placed about 50 *cm* away from the rotational axis of the turntable. Ideally the optical axis of the camera for acquiring object's silhouettes lies nearly in the rotational plane of the turntable, orthogonal to the rotational axis. The camera for acquiring the projection of the laser plane onto the object views the turntable from an angle of about  $45^\circ$ . The laser is directed such that the light plane projected contains the rotational axis of the turntable. Prior to any acquisition, the system is calibrated in order to determine the inner and outer orientation of the cameras and the rotational axis of the turntable (for details see [14]).

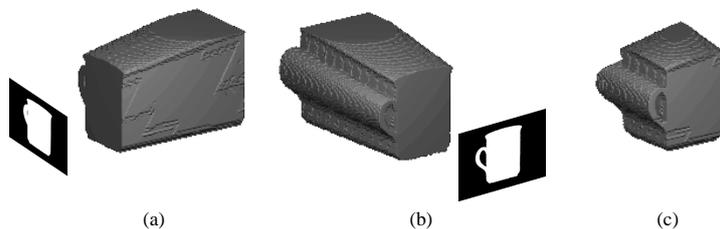


**Figure 1: Acquisition system.**

### 3 Combination of Algorithms

An input image for SfS defines a conic volume in space which contains the object to be modeled (Figure 2a). Another input image taken from a different view defines another conic volume containing the object (Figure 2b). Intersection of the two conic volumes narrows down the space the object can possibly occupy (Figure 2c). With an increasing number of views the intersection of all conic volumes approximates the actual volume occupied by the object better and better, converging to the 3d visual hull of the object [8]. Therefore by its nature SfS defines a *volumetric* model of an object.

An input image for SfL using laser light defines solely the points on the surface of the object which intersect the laser plane (Figure 3a). Multiple views provide a cloud of points belonging to the object surface (Figure 3b), which is a *surface* model of the object.



**Figure 2: Two conic volumes and their intersection.**

The main problem to be addressed in an attempt to combine these two methods is how to adapt the two representations to one another, i.e. how to build a common 3d model representation. One possibility would be to build a separate SfL surface model and a SfS volumetric model followed by converting one model to the other and intersecting them. But if we want to estimate the volume of an



**Figure 3: Laser projection and cloud of points.**

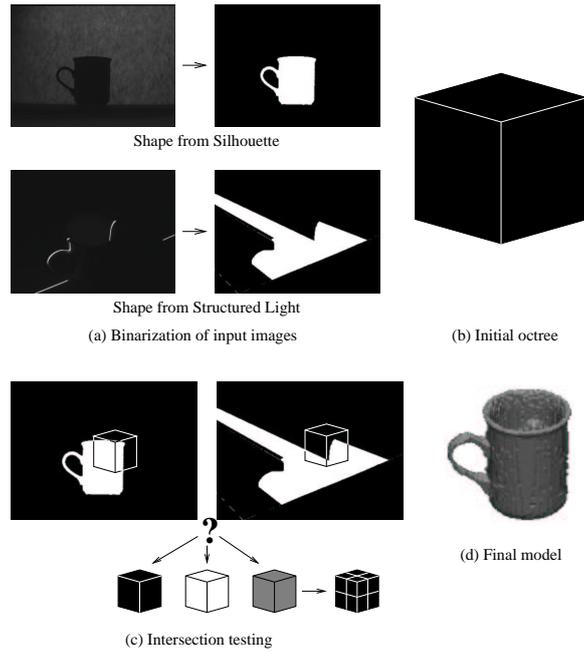
object using our model, any intermediate surface models should be avoided because of the problems of conversion to a volumetric model. Therefore, our approach proposes building a single volumetric model from the ground up, using both underlying methods (see Figure 4):

The first step between the image acquisition and creation of the final 3d model of an object consists of converting the images acquired into binary images. A pixel in such a binary image should have the value 0 if it represents a point in 3d space which does not belong to the object *for sure*, and the value of 1 otherwise. The binarization is performed on input images for both SfS and SfL.

For the SfS part of the method presented, a reliable extraction of the object's silhouette from an acquired image is of crucial importance for obtaining an accurate 3d model of an object. In addition to the images of the object (Figure 4a, upper image) taken from different viewpoints, an image of the acquisition space is taken, without any object in it. Then, the absolute difference between this image and an input image is built, which creates an image with a uniform background and a high contrast between the object and the background.

An input image for SfL contains the projection of a laser plane onto the object (Figure 4a, lower image). A white pixel in this image represents a 3d point on the object's surface which intersects the laser plane. A black pixel represents a 3d point in the laser plane which does not belong to the object's surface — it is either inside the object or it does not belong to the object at all. Based on the known position of the laser, an input image (Figure 4a, lower left image) is converted to an image approximating the intersection of the laser plane with the whole object (Figure 4a, lower right image).

Our approach builds a 3d model of an object performing the following steps (illustrated in Figure 4): First, both of the input images (SfS and SfL) are binarized such that the white image pixels *possibly* belong to the object and the black pixels *for sure* belong to the background (Figure 4a). Then, the initial octree containing one single root node marked "black" is built (Figure 4b). Black nodes are subsequently checked by projecting the nodes into all SfS binarized input images and intersecting them with the image silhouettes of the object (Figure 4c). As the result of the intersection the node can remain "black" (if it lies within the object) or set to "white" (it lies outside the object) or "grey" (it lies partly within and partly outside the object). If the resulting node is not white, it is projected



**Figure 4: Algorithm overview.**

into the binarized SfL image representing the nearest laser plane to the node and again intersected. All grey nodes are divided into 8 child nodes all of which are marked "black" and the intersection test is performed in each of the black nodes. This subdivision of grey nodes is done until there are no grey nodes left or a subdivision is not possible (voxel size), which results in the final model (Figure 4d).

## 4 Results

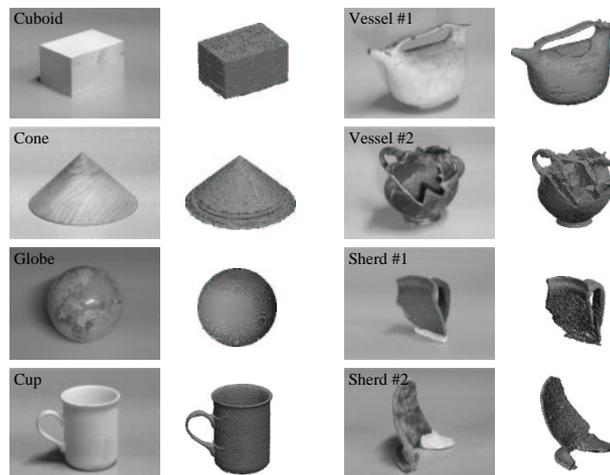
For tests with synthetic objects we can build a model of a virtual camera and laser and create input images in a way that the images fit perfectly into the camera model. We assume having a virtual camera with focal length  $f = 20 \text{ mm}$ , placed on the  $y$  axis of the world coordinate system,  $2000 \text{ mm}$  away from its origin. We set the distance between two sensor elements of the camera to  $d_x = d_y = 0.01 \text{ mm}$ . The laser is located on the  $z$  axis of the world coordinate system,  $850 \text{ mm}$  away from its origin, and the turntable  $250 \text{ mm}$  below the  $x$ - $y$  plane of the world coordinate system, with its rotational axis identical to the  $z$  world axis. We build input images with size  $640 \times 480$  pixels, in which 1 pixel corresponds to  $1 \text{ mm}$  in the  $x$ - $z$  plane of the world coordinate system. As the synthetic object we create a sphere with radius  $r = 200 \text{ mm}$ . Since the sphere does not contain any cavities, SfS can also reconstruct it completely. Therefore, we can measure the accuracy of each of the methods independently, as well as of the combined method.

In a test we build models using 360 views with the constant angle of  $1^\circ$  between two views, while increasing octree resolution. It turned out that the SfS method performed best with an octree reso-

lution of  $128^3$ , where the approximation error was +0.83% of the actual volume, the structured light method with a resolution of  $256^3$  and +0.29% error (the other method produced there an error of -1.42%). In the second test we build models with constant octree resolution of  $256^3$  and increasing number of views. Regarding the number of views, there was no significant difference between the two methods. Using 20 instead of 360 views was sufficient for both methods to create models less than 1% different from the models built using 360 views.

For tests with real objects we used 8 objects: a metal cuboid, a wooden cone, a globe, a coffee cup, two archaeological vessels and two archaeological sherds. The real volume of the first 3 objects can be computed analytically. For the two vessel it could be theoretically measured by putting water into the objects, but it has not been done since the vessels do have holes, which we are not allowed to close, so for these objects we can only compare the bounding cuboid of the model and the object. Figure 5 shows the objects and their models built using 360 views for each of the underlying methods and the octree resolution  $256^3$ .

The error of the computed volume for real objects was between 3% and 13%, by an order of magnitude larger than the errors with synthetic objects. The main reason turned out to be the threshold based binarization of silhouette images, which interpreted parts of the object as the background, especially close to the turntable surface. That explains why the error was the biggest for the cone and the smallest for the globe (see Table 1). The cone has a large base leaning on the turntable, while the globe only touches the turntable in an almost tangential way.



**Figure 5: Real objects and their models.**

## 5 Conclusion

In this paper a combination of a SfS method with a SfL method was presented, which creates a 3d model of an object from images of the object taken from different viewpoints. The algorithm employs only simple matrix operations for all the transformations and it is fast, because even for

<i>object</i>	<i>octree</i>	<i>#views</i>	<i>volume</i>	<i>vol.error</i>
synth. sphere	—	analytic	33 510 322	—
	64 <sup>3</sup>	360+360	35 241 984	+5.17%
	128 <sup>3</sup>	360+360	33 786 880	+0.83%
	256 <sup>3</sup>	360+360	33 034 528	-1.42%
	256 <sup>3</sup>	180+180	33 067 552	-1.32%
	256 <sup>3</sup>	20+20	33 230 464	-0.83%
synth. cuboid	—	analytic	420 000	—
	64 <sup>3</sup>	360+360	432 000	+2.86%
	128 <sup>3</sup>	360+360	420 000	0.00%
	256 <sup>3</sup>	360+360	420 000	0.00%
	256 <sup>3</sup>	180+180	426 071	+1.45%
	256 <sup>3</sup>	20+20	435 402	+3.67%
real cuboid	—	analytic	420 000	—
	256 <sup>3</sup>	360+360	384 678	-8.41%
cone	—	analytic	496 950	—
	256 <sup>3</sup>	360+360	435 180	-12.43%
globe	—	analytic	1 756 564	—
	256 <sup>3</sup>	360+360	1 717 624	-2.22%
cup	—	analytic	N/A	—
	256 <sup>3</sup>	360+360	276 440	N/A
vessel #1	—	analytic	N/A	—
	256 <sup>3</sup>	360+360	336 131	N/A
vessel #2	—	analytic	N/A	—
	256 <sup>3</sup>	360+360	263 696	N/A
sherd #1	—	analytic	N/A	—
	256 <sup>3</sup>	360+360	35 911	N/A
sherd #2	—	analytic	N/A	—
	256 <sup>3</sup>	360+360	38 586	N/A

**Table 1: Volume of objects and their models.**

highly detailed objects, a high resolution octree (256<sup>3</sup> voxels) and a high number of input views (36), the computational time hardly exceeds 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.

## References

- [1] H. Baker. Three-dimensional modelling. In *Proc. of 5th Intl. Conf. on AI*, pages 649–655, 1977.
- [2] B.G. Baumgart. *Geometric Modeling for Computer Vision*. PhD thesis, Stanford AI, 1974.
- [3] P.J. Besl. Active, optical range imaging sensors. *MVA*, 1(2):127–152, 1988.
- [4] F.W. DePiero and M.M. Trivedi. 3-D computer vision using structured light: Design, calibration, and implementation issues. *Advances in Computers*, 43:243–278, 1996.
- [5] R. M. Haralick and L. G. Shapiro. Glossary of computer vision terms. *PR*, 24(1):69–93, 1991.

- [6] M. Kampel and R. Sablatnig. Automated 3d recording of archaeological pottery. In D. Bearman and F. Garzotto, editors, *Proc. of Intl. Conf. on Cultural Heritage and Technologies in the 3rd Millennium*, pages 169–182, 2001.
- [7] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. *IJCV*, 38(3):197–216, 2000.
- [8] A. Laurentini. The visual hull concept for silhouette-based image understanding. *PAMI*, 16(2):150–162, 1994.
- [9] C. Liska and R. Sablatnig. Estimating the next sensor position based on surface characteristics. In *ICPR00*, volume I, pages 538–541, 2000.
- [10] W. N. Martin and J. K. Aggarwal. Volumetric description of objects from multiple views. *PAMI*, 5(2):150–158, 1983.
- [11] W. Niem. Error analysis for silhouette-based 3D shape estimation from multiple views. In N. Sarris and M.G. Strintzis, editors, *Proc. of Intl. Workshop on Synthetic-Natural Hybrid Coding and 3D Imaging*, pages 143–146, 1997.
- [12] M. Potmesil. Generating octree models of 3D objects from their silhouettes in a sequence of images. *CVGIP*, 40:1–29, 1987.
- [13] R. Szeliski. Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58(1):23–32, 1993.
- [14] S. Tosovic and R. Sablatnig. 3d modeling of archaeological vessels using shape from silhouette. In *Proc. of Conf. on 3-D Digital Imaging and Modeling*, pages 51–58, 2001.