Octree-based Fusion of Shape from Silhouette and Shape from Structured Light *

Martin Kampel, Srdan Tosovic, and Robert Sablatnig
Vienna University of Technology,
Institute of Computer Aided Automation,
Pattern Recognition and Image Processing Group
Favoritenstr. 9, 183-2, A-1040 Vienna
{kampel,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 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 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* (SfSL) method presented in this paper was performed within the *Computer Aided Classification of Ceramics* [7] project, which aims to provide an objective and automated method for classification and reconstruction of archaeological pottery. 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 [13]. 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 interest to archaeologists too, since the volume estimation allows also a more precise classification [13].

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 [16, 15]. 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* [8].

There have been many works on construction of 3D models of objects from multiple views ([1, 10, 4, 18, 15]). Szeliski [16] first creates a low resolution octree model quickly and then refines this model iteratively, by intersecting each new silhouette with the already existing model. Niem [12] uses pillar-like volume elements instead of an octree for the model representation. De Bonet and Viola [3] extended the idea of voxel reconstruction to transparent objects by introducing the Roxel algorithm — a responsibility weighted 3D volume reconstruction. Wong and Cipolla [20] use uncalibrated silhouette images and recover the camera positions and orientations from circular motions.

SfS can be applied on 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. This condition is very hard to hold since most of the archaeological vessels do have concavities. To discover these concavities we use SfSL, which is based on active triangulation [2, 6]. Most laser light based SfSL methods use a camera, a calibrated laser ray or plane and a motion platform — usually a linear slide or a turntable.

The work of Szeliski [16] was used as a basis for the SfS and the work of Liska [9] as a basis for the SfSL approach presented in this paper which is organized as follows. Section 2 describes the equipment used for acquisition. Section

*This work was partly supported by the Austrian Science Foundation (FWF) under grant P13385-INF, the European Union under grant IST-1999-20273 and the Austrian Federal Ministry of Education, Science and Culture.

Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)
0-7695-1521-5/02 $17.00 © 2002 IEEE

3 presents the combination strategy proposed. Experimental results with both synthetic and real data are given in Section 4. At the end of the paper conclusions are drawn and future work is outlined.

2 Acquisition System

The acquisition system, shown in Figure 1, consists of a turntable (diameter 50 cm), two monochrome CCD-cameras (f=16 mm, 768×576 pixels), a laser and a lamp. Both cameras are placed in a distance of about 50 cm from the rotational axis of the turntable. Ideally the optical axis of the camera for acquiring object's silhouettes (Camera-1 in Figure 1) 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 (Camera-2 in Figure 1) views the turntable in an angle of about $45°$. The laser is directed such that the light plane it projects contains the rotational axis of the turntable. Camera-2 from Figure 1 views the light plane also from an angle of about $45°$. The relative position of the two cameras to one another is not important, since the acquisition of the silhouettes and the acquisition of the laser light projection are independent from one another.

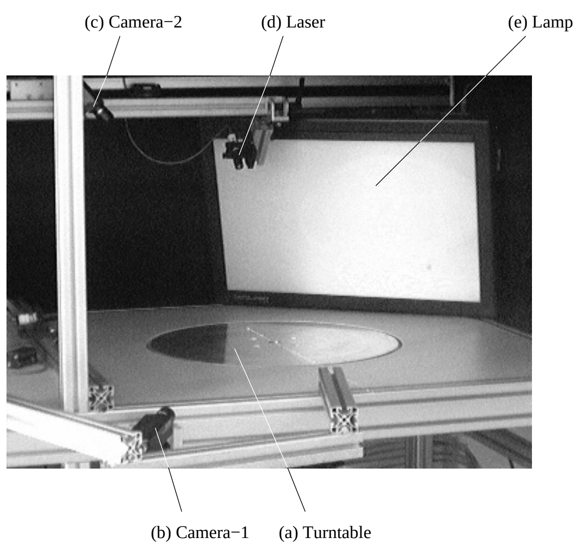

Figure 1. Acquisition system

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. We used the calibration technique proposed by Roger Y. Tsai [17], for several reasons: it is efficient and accurate, lens distortion can be taken into account but also ignored if desired, and there is a publicly available implementation [19]. In our experiments, the average calibration error was 0.5 pixel or less (measured in the image plane), which is sufficient for our approach, because the smallest unit processed in an image is 1 pixel.

3 Fusion of Algorithms

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 SfSL.

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. If the background brightness is not uniform, the silhouette extraction can be a difficult task. For that reason, in addition to the images of the object (Figure 2a, 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. Next, thresholding is used to create a binary image (Figure 2a, upper image) where pixels with the value 1 represent the object's silhouette and those with value 0 the background.

Another option for extracting object silhouettes from input images would be to use edge detection [11] instead of thresholding. This approach could be more accurate, even a sub-pixel precision could be reached, but it is also more complex.

An input image for SfSL contains the projection of a laser plane onto the object (Figure 2a, 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. The creation of a SfSL binary image is more complex. Based on the known position of the laser, an input image (Figure 2a, lower left image) is converted to an image approximating intersection of the laser plane with the whole object (Figure 2a, lower right image).

Our approach builds a 3D model of an object performing the following steps (illustrated in Figure 2): First, both of the input images (SfS and SfSL) are binarized such that the white image pixels *possibly* belong to the object and the black pixels *for sure* belong to the background (Figure 2a). Then, the initial octree containing one single root node marked "black" is build (Figure 2b). Black nodes are subsequently checked by projecting the nodes into all SfS binarized input images and intersecting them with the im-

Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)
0-7695-1521-5/02 $17.00 © 2002 **IEEE**

age silhouettes of the object (Figure 2c). As the result of the intersection the node can remain ”black” (if it lies within the object) or be 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 into the binarized SfSL 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 2d).

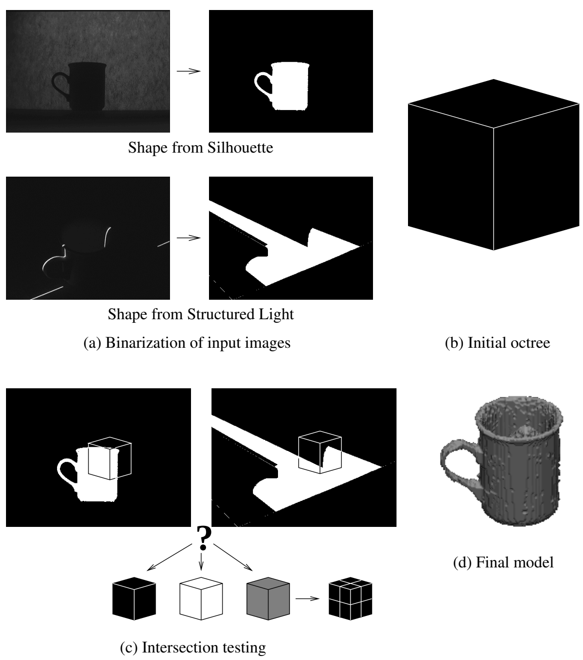

Figure 2. Algorithm overview

4 Results

Experiments were performed with both synthetic and real objects. For synthetic objects we built a model of a virtual camera and laser and created input images such that the images fit perfectly into the camera model. For both synthetic and real objects we compare the volume and the size of the bounding cuboid of the model with the volume and size of the bounding cuboid of the object. As synthetic objects we created a virtual sphere with the radius $200\ mm$, and a virtual cuboid with dimensions $100 \times 70 \times 60\ mm$. The images of the sphere were constructed such that both SfS and SfSL alone can reconstruct the object completely, whereas for the cuboid a more realistic case was simulated, where the structured light images contain occlusions. The models of these objects were constructed with different parameter values, such as the number of views used and the maximal octree resolution.

The tests with the sphere showed that SfS and SfSL perform similarly when they have perfect input images — starting from resolution 128^3, both methods were able to create models with the approximation error of 2% or less. Regarding the number of views, 20 views were sufficient for both methods in order to create models with the volume less than 1% different from the models built using 360 views. With the synthetic cuboid, neither of the methods was able to reconstruct the cuboid completely, but the combined method constructed its perfect model starting from the resolution 128^3. However, even if using 180 views instead of 360, the volume error of the cuboid was greater than 1% (1.45%), which indicates that flat surfaces are more difficult to model with our method. Table 1 summarizes the results of the models built.

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 3 shows the objects and their models built using 360 views for each of the underlying methods and the octree resolution 256^3.

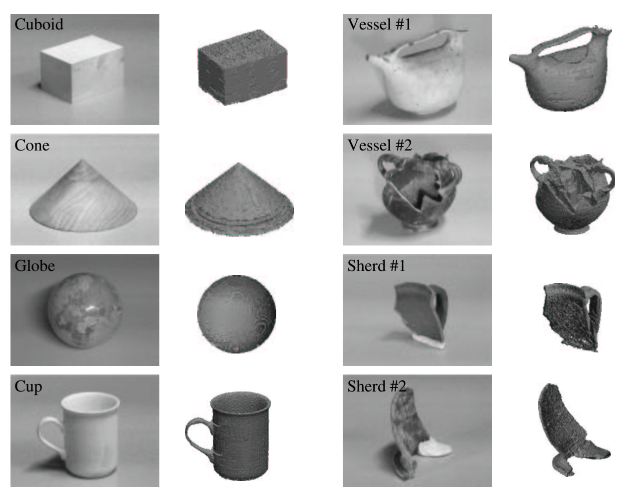

Figure 3. Real objects and their models

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 largest 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

Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)
0-7695-1521-5/02 $17.00 © 2002 IEEE

the turntable in an almost tangential way.

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

Table 1. Volume of objects and their models

5 Conclusion

This paper presented a 3D modeling method based on combination of SfS and laser based SfSL, using a turntable to obtain multiple views of an object. The purpose of combining SfS and SfSL was to create a method which will use the advantages and overcome the weaknesses of both underlying methods and create complete models of arbitrarily shaped objects. The experiments with synthetic objects showed that construction of nearly perfect models is possible, limited only by image and model resolution. In the experiments with real objects the results were less accurate, but the algorithm was able to produce complete and visually faithful models for all objects, including sherds and vessels with concave surfaces and a handle.

Overall, our combined modeling approach proved to be useful for automatic creation of models of arbitrarily shaped objects. With respect to its archaeological application it can provide models of any kind of archaeological pottery. Furthermore, the volume of an object can be estimated, including the inside volume of objects such as bowls or cups. However, for high precision measurements of the volume our method did not produce highly accurate results, but it gave a good rough estimate, which is sufficient for most archaeological applications. Higher accuracy could be achieved by improving the binarization of input images, which showed to be the main reason for relatively large errors for real objects. A possible enhancement to our method would be to take additional color images of an object and perform texture mapping onto the model, which would improve the visual impression of the models built.

References

[1] H. Baker. Three-dimensional modelling. In *Proc. of IJCAI*, pages 649–655, 1977.

[2] P. Besl. Active, optical range imaging sensors. *MVA*, 1(2):127–152, 1988.

[3] J. S. D. Bonet and P. Viola. Roxels: Responsibility weighted 3D volume reconstruction. In *Proc. of 7th ICCV*, pages 418–425, 1999.

[4] C. H. Chien and J. K. Aggarwal. Volume/surface octrees for the representation of three-dimensional objects. *CVGIP*, 36:100–113, 1983.

[5] J. Davis and X. Chen. A laser range scanner designed for minimum calibration complexity. In *Proc. of 3rd Intl. Conf. on 3-D Digital Imaging and Modeling*, pages 91–98, 2001.

[6] F. DePiero and M. Trivedi. 3-D computer vision using structured light: Design, calibration, and implementation issues. *Advances in Computers*, 43:243–278, 1996.

[7] M. Kampel and R. Sablatnig. Automated 3D recording of archaeological pottery. In *D. Bearman and F. Garzotto (Ed.). Proc. of the Int. Conf on Cultural Heritage and Technologies in the Third Millennium Milan, Italy*, pages 169–182, 2001.

[8] K. Kutulakos and S. Seitz. A theory of shape by space carving. *IJCV*, 38(3):197–216, July 2000.

[9] C. Liska and R. Sablatnig. Estimating the next sensor position based on surface characteristics. In *Proc. of ICPR*, volume I, pages 538–541, 2000.

[10] W. N. Martin and J. K. Aggarwal. Volumetric description of objects from multiple views. *IEEE-PAMI*, 5(2):150–158, 1983.

[11] V. S. Nalwa and T. O. Binford. On detecting edges. *IEEE-PAMI*, 8:699–714, November 1986.

[12] W. Niem. Robust and fast modelling of 3D natural objects from multiple views. In *Image and Video Processing II, Proc. of SPIE*, pages 388–397, 1994.

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

[14] J. Park, G. N. DeSouza, and A. C. Kak. Dual-beam structured-light scanning for 3-D object modeling. In *Proc. of 3rd Intl. Conf. on 3-D Digital Imaging and Modeling*, pages 65–72, 2001.

[15] M. Potmesil. Generating octree models of 3D objects from their silhouettes in a sequence of images. *CVGIP*, 40:1–29, 1987.

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

[17] R. Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. of CVPR*, pages 364–374, 1986.

[18] J. Veenstra and N. Ahuja. Efficient octree generation from silhouettes. In *Proc. of CVPR*, pages 537–542, 1986.

[19] R. G. Willson. Tsai camera calibration software. http://www.cs.cmu.edu/˜rgw/TsaiCode.html.

[20] K. Y. K. Wong and R. Cipolla. Structure and motion from silhouettes. In *Proc. of ICCV*, pages 217–222, 2001.

0-7695-1521-5/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY