On Combining Shape from Silhouette and Shape from Structured Light *

Srdan Tosovic, Robert Sablatnig, and Martin Kampel
Vienna University of Technology,
Institute of Computer Aided Automation,
Pattern Recognition and Image Processing Group
Favoritenstr. 9, 183-2, A-1040 Vienna
e-mail: {tos,sab,kampel}@prip.tuwien.ac.at

Abstract

This paper presents an octree based method of three-dimensional reconstruction of objects using a combination of two different methods, Shape from Silhouette and Shape from Structured Light, focusing on reconstruction of archaeological vessels. Shape from Silhouette is a method suitable for reconstruction of objects with handles, whereas it is unable to reconstruct concavities on an object's surface, such as inside of a bowl. Shape from Structured Light can reconstruct such concavities, but it often creates incomplete models because of camera and light occlusions. The purpose of combining these two methods is to overcome the weaknesses of one method through the strengths of the other, making it possible to construct complete models of arbitrarily shaped objects. The construction is based on multiple views of an object using a turntable in front of stationary cameras. Results of the algorithm developed are presented for both synthetic and real objects.

1 Introduction

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 which the object's silhouette represents the only interesting feature of the image [21, 18]. 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* [12].

Shape from Silhouette can be applied on objects with variety of shapes, including objects with handles, like many archaeological vessels and sherds. However, concavities on an object's

*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.

surface remain invisible for this method, making it unusable for reconstruction of the inside of a bowl or a cup or the inner side of a sherd. Therefore, another method, Shape from Structured Light, is used to discover the concavities.

Shape from Structured Light is a method which constructs a surface model of an object based on projecting a sequence of well defined light patterns onto the object. The patterns can be in the form of coded light stripes [11] or a ray or plane of laser light [13]. The 3D coordinates of the points on the object's surface are recovered using active triangulation [3, 9].

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 [14] constructed volume segment models from orthographic projection of silhouettes. Chien and Aggarwal [7] constructed an object's octree model from its three orthographic projections. Veenstra and Ahuja [23] extended this approach to thirteen standard orthographic views. Potmesil [18] created octree models using arbitrary views and perspective projection. For each of the views he constructs an octree representing the corresponding conic volume and then intersects all octrees. In contrast to this, Szeliski [21] first creates a low resolution octree model quickly and then refines this model interactively, 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 [20] in contrast, perform the intersections in 3D-space. Niem [15] uses pillar-like volume elements (pillars) rather than octree for model representation. De Bonet and Viola [4] extended the idea of voxel reconstruction to transparent objects by introducing the Roxel algorithm — a responsibility weighted 3D volume reconstruction. Wong and Cipolla [26] use uncalibrated silhouette images and recover the camera positions and orientations from circular motions.

Most laser light based Shape from Structured Light methods use a camera, a calibrated laser ray or plane and a motion platform — usually a linear slide or a turntable. Borgese et al. [5] use a pair of standard video cameras, a laser pointer, and a special hardware that lets the laser spot be detected with high reliability and accuracy. By obtaining a single surface point at each step, this method implies a slow, sparse sampling of the surface. Liska [13] uses two lasers aligned to project the same plane, a camera and a turntable. Using two lasers eliminates some of the light occlusions but not the camera occlusions, resulting in incomplete models for many objects. Park et al. [17] built a DSLS (Dual Beam Structured Light) scanner, which uses a camera mounted on a linear slide and two non-overlapping laser planes, resulting in denser range images. Davis and Chen [8] use two calibrated fixed cameras viewing a static scene and an uncalibrated laser plane which is freely swept over the object.

The work of Szeliski [21] was used as a basis for the Shape from Silhouette and the work of Liska [13] as a basis for the Shape from Structured Light approach presented in this paper.

The paper is organized as follows. Section 2 describes the equipment used for acquisition. Section 3 describes the octree model representation and Section 4 presents the combination strategy proposed. Experimental results with both synthetic and real data are given in Section 5 and at the end of the paper conclusions are drawn and future work is outlined.

2 Acquisition System

The acquisition system consists of the following devices:

- a turntable (Figure 1a) with a diameter of 50 cm, whose desired position can be specified with an accuracy of $0.05°$ (however, the minimal relative rotation angle is $1.00°$).
- 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 the scene for the acquisition of the silhouette of the object [10]. The object should be clearly distinguishable from the background independent from the object's shape or the type of its surface.

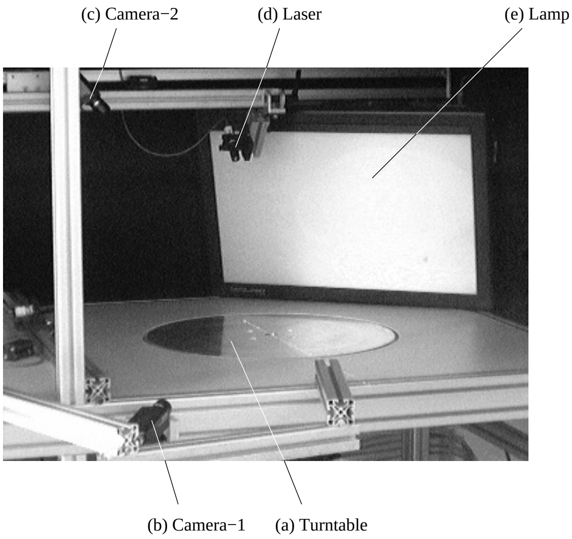

Figure 1: Acquisition System

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 $45°$. The laser is directed in such that the light plane it projects contains the rotational axis of the turntable. The second camera 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.

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 [22], 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 on Internet [25]. 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 Octree Model Representation

There are many different model representations in computer vision and computer graphics used. Here we will mention only the most important ones. Surface-based representations describe the surface of an object as a set of simple approximating patches, like planar or quadratic patches [2]. Generalized cylinder representation [19] defines a volume by a curved axis and a cross-section function at each point of the axis. Overlapping sphere representation [16] describes a volume as a set of arbitrarily located and sized spheres. Approaches such as these are efficient in representing a specific set of shapes but they are not flexible enough to describe arbitrary solid objects. Two of the most commonly used representations for solid volumes are boundary representation (B-Rep) [24] and constructive solid geometry (CSG) [24, 19].

An octree [6] 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

An octree as described above contains binary information in the leaf nodes and therefore it is called a binary octree, and it is suitable for representation of 3D objects where the shape of the object is the only object property that needs to be modeled by the octree. Non-binary octrees can contain other information in the leaf nodes, e.g., the cube color in RGB-space. For the 3D modeling approach presented in this work, a binary octree model is sufficient to represent 3D objects.

The octree representation has several advantages [6]: 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 depend on the maximal octree depth and therefore cannot have smooth surfaces. However, this is a problem with any kind of voxel-based volumetric representation.

4 Fusion of the Algorithms

As noted in Section 1, Shape from Silhouette defines a *volumetric* model of an object, whereas Shape from Structured Light defines a *surface* model of an object. The main problem that needs 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. This can be done in several ways:

- Build the *Shape from Silhouette*'s volumetric model and the *Shape from Structured Light*'s surface model independently from one another. Then, either convert the volumetric model to a surface model and use a combination of the two surface models to create the final representation or convert the surface model to a volumetric model and use a combination of the two volumetric models to create the final representation. Depending on the properties of the two models (e.g., whether they represent a subset or a superset of the object), their combination can mean their union or their intersection or some more complex operation.
- Use a common 3D model representation from the ground up, avoiding any model conversions. That means either design a volume based Shape from Structured Light algorithm or a surface based Shape from Silhouette algorithm.

Generally, the conversion of a surface model to a volumetric model is a complex task, because if the surface is not completely closed, it is hard to say whether a certain voxel lies inside or outside the object. With closed surfaces one could follow a line in 3D space starting from the voxel observed and going in any direction and count how many times the line intersects the surface. For an odd number of intersections one can say that the voxel belongs to the object. But even in this case there would be many special cases to handle, e.g. when the chosen line is tangential to the object's surface.

This reasoning lead us to the following conclusions:

- Building a separate Shape from Structured Light surface model and a Shape from Silhouette volumetric model followed by converting one model to the other and then combining them is mathematically complex and computationally costly.
- If we want to estimate the volume of an 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 in each step (illustrated in Figure 2):

1. Binarize the acquired images for both Shape from Silhouette and Shape from Structured Light in such a way that the white image pixels *possibly* belong to the object and the black pixels *for sure* belong to the background (see Figure 2a). A silhouette binary image is created by extraction of the object's silhouette through simple thresholding of the image. The creation of a structured light binary image is more complex. Based on the known position of the laser, an input image (representing the intersection of the laser plane with the object's surface) is converted to an image approximating intersection of the laser plane with the whole object.
2. Build the initial octree, containing one single root node marked ”black”. (Figure 2b). This node is said to be at the level 0.

3. All black nodes of the `current level` are assumed to be in a linked list. If there are no nodes in the `current level`, the final model has been build so jump to Step 8. Otherwise, continue with Step 4.
4. Project the `current node` of the `current level` into all Shape from Silhouette binary images and intersect it with the image silhouettes of the object. 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 "gray" (it lies partly within and partly outside the object), see image on the left in Figure 2c. Note that the meaning of "black" in the octree and in the binary images is inverted — a node is black if it's projection lies lies entirely in the white area of an image.
5. If the `current node` after Step 4 is not white, it is projected into the Shape from Structured Light binary image representing the nearest laser plane to the node (ideally the plane intersecting the node center) and intersected with the area representing the intersection of the object and the laser plane (image on the right in Figure 2c). Other structured light images, representing planes which do not intersect the `current node`, are irrelevant for determination of its color.
6. If the node is set to Gray it is divided into 8 child nodes of the `current level` $+ 1$, all of which are marked "black"
7. Processing of the `current node` is finished. If there are more nodes in the `current level` set the `current node` to the next node and go back to Step 4. If all nodes of the `current level` have been processed, increment the `current level` and go to Step 3.
8. The final octree model has been built (Figure 2d).

5 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 in such a way that the images fit perfectly into the camera model. Doing so the accuracy of the models constructed can be analyzed, without impact of camera calibration errors. 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 in such a way, that both Shape from Silhouette and Shape from Structured Light 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. Figure 3 shows the models built using 360 silhouette and 360 structured light views, with the constant angle of 1° between two views, and the octree resolution 256^3. The tests with the

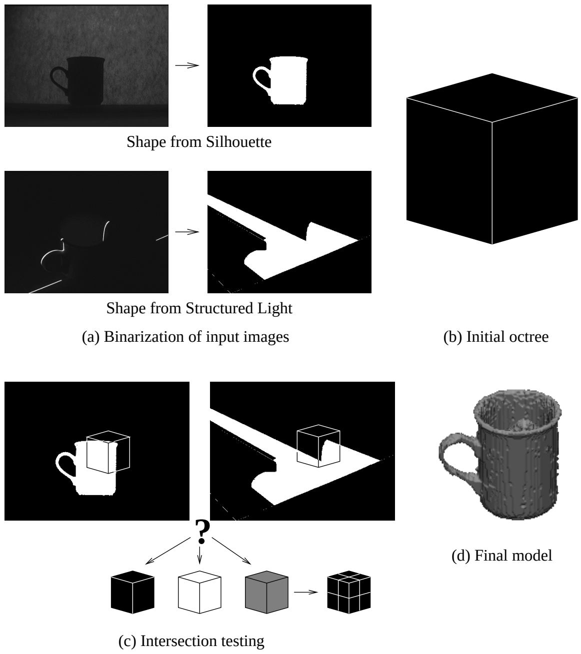

Figure 2: Algorithm overview

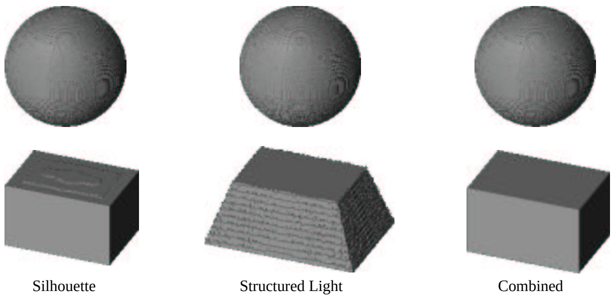

Figure 3: 3D models of synthetic sphere and cuboid

sphere showed that Shape from Silhouette and Shape from Structured Light 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 was 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 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 remaining 5 objects it can be measured by putting the objects in the water, but it has not been done yet at the time of writing this paper, so for these objects we can only compare the bounding cuboid of the model and the object. Figure 4 shows the objects and their models built using 360 views for each of the underlying methods and the octree resolution 256^3. In addition, Figure 5 compares the models of the cup and one of the sherds built using one of the methods only and the combined method, illustrating the necessity of using both methods in order to construct complete models of objects.

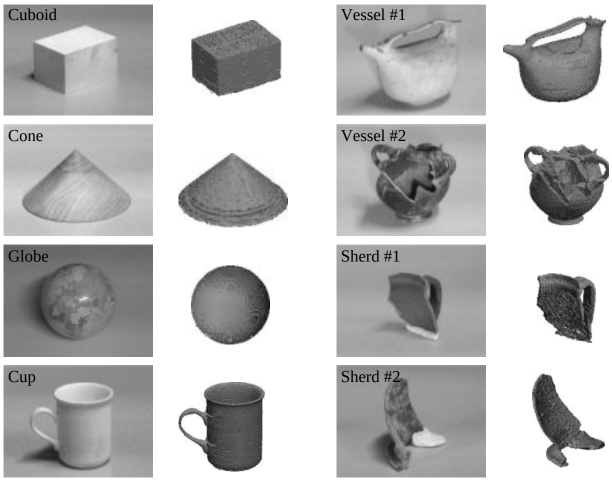

Figure 4: Real objects and their models

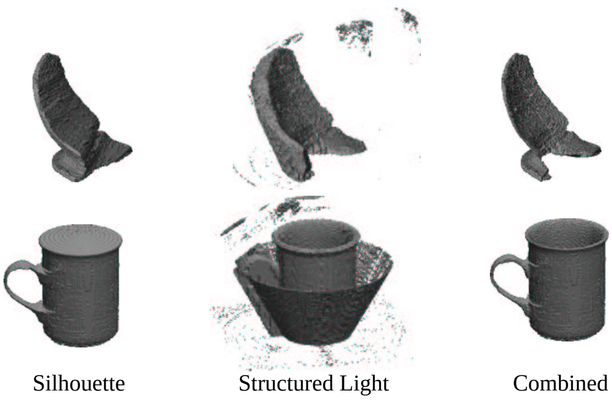

Figure 5: Silhouette and structured light based models

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 the turntable in an almost tangential way.

The dimensions and the volume of the objects presented in this section and their 3D models are summarized in Table 1.

object	*octree*	*#views*	*dimensions (mm)*	*volume (mm^3)*	*volume error*
synth. sphere	—	analytic	$400.0 \times 400.0 \times 400.0$	33 510 322	—
	64^3	360+360	$400.0 \times 400.0 \times 400.0$	35 241 984	+5.17%
	128^3	360+360	$400.0 \times 400.0 \times 400.0$	33 786 880	+0.83%
	256^3	360+360	$396.0 \times 396.0 \times 400.0$	33 034 528	-1.42%
	256^3	180+180	$396.0 \times 396.0 \times 400.0$	33 067 552	-1.32%
	256^3	20+20	$400.0 \times 400.0 \times 400.0$	33 230 464	-0.83%
synth. cuboid	—	analytic	$100.0 \times 70.0 \times 60.0$	420 000	—
	64^3	360+360	$100.0 \times 70.0 \times 60.0$	432 000	+2.86%
	128^3	360+360	$100.0 \times 70.0 \times 60.0$	420 000	0.00%
	256^3	360+360	$100.0 \times 70.0 \times 60.0$	420 000	0.00%
	256^3	180+180	$100.0 \times 72.0 \times 60.0$	426 071	+1.45%
	256^3	20+20	$104.0 \times 73.0 \times 60.0$	435 402	+3.67%
real cuboid	—	analytic	$100.0 \times 70.0 \times 60.0$	420 000	—
	256^3	360+360	$101.0 \times 71.0 \times 60.0$	384 678	-8.41%
cone	—	analytic	$156.0 \times 156.0 \times 78.0$	496 950	—
	256^3	360+360	$150.1 \times 149.4 \times 77.5$	435 180	-12.43%
globe	—	analytic	$149.7 \times 149.7 \times 149.7$	1 756 564	—
	256^3	360+360	$149.1 \times 148.2 \times 144.6$	1 717 624	-2.22%
cup	—	analytic	$113.3 \times 80.0 \times 98.9$	N/A	—
	256^3	360+360	$111.6 \times 79.0 \times 98.3$	276 440	N/A
vessel #1	—	analytic	$141.2 \times 84.8 \times 93.7$	N/A	—
	256^3	360+360	$139.2 \times 83.2 \times 91.4$	336 131	N/A
vessel #2	—	analytic	$114.2 \times 114.6 \times 87.4$	N/A	—
	256^3	360+360	$113.0 \times 111.9 \times 86.4$	263 696	N/A
sherd #1	—	analytic	$51.8 \times 67.0 \times 82.2$	N/A	—
	256^3	360+360	$51.0 \times 66.0 \times 79.4$	35 911	N/A
sherd #2	—	analytic	$76.0 \times 107.3 \times 88.5$	N/A	—
	256^3	360+360	$74.9 \times 103.9 \times 86.2$	38 586	N/A

Table 1: Dimensions and volume of objects and their models

6 Conclusion

This paper presented a 3D modeling method based on combination of Shape from Silhouette and laser based Shape from Structured Light, using a turntable to obtain multiple views of an object. The purpose of combining Shape from Silhouette and Shape from Structured Light 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. Only the inside of deep objects could not be completely recovered, due to camera and light occlusions.

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. The models can also be intersected with arbitrary planes, resulting in profile sections of a sherd or a vessel. 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 5th International Joint Conference on Artificial Intelligence*, pages 649–655, 1977.

[2] P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:167–192, March 1988.

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

[4] J. S. De Bonet and P. Viola. Roxels: Responsibility weighted 3D volume reconstruction. In *Proc. of 7th IEEE International Conference on Computer Vision*, pages 418–425, 1999.

[5] N. A. Borghese, G. Ferrigno, G. Baroni, A. Pedotti, S. Ferrari, and R. Savarè. Autoscan: A flexible and portable 3D scanner. *IEEE Computer Graphics and Applications*, 18(3):38–41, 1998.

[6] 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.

[7] 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.

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

[9] 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.

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

[11] M. Kampel and R. Sablatnig. Range image registration of rotationally symmetric objects. In N. Brändle, editor, *Proc. of Computer Vision Winter Workshop*, pages 69–77, 1999.

[12] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):197–216, July 2000.

[13] C. Liska and R. Sablatnig. Estimating the next sensor position based on surface characteristics. In *Proc. of 15th. Intl. Conf. on Pattern Recognition, Barcelona, Spain*, pages Vol I: 538–541, 2000.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

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

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

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

[22] R. Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, 1986.

[23] 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.

[24] A. Watt. *3D Computer Graphics*. Addison Wesley, 3 edition, December 1999.

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

[26] K. Y. K. Wong and R. Cipolla. Structure and motion from silhouettes. In *Proc. of 8th IEEE International Conference on Computer Vision*, pages 217–222, 2001.