A Next View Planning Technique for Shape from Silhouette and Shape from Structured Light[1)]

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

Abstract:
In order to create a complete three-dimensional model of an object based on its two-dimensional images, the images have to be acquired from different views. An increasing number of views generally improves the accuracy of the final 3D model but it also increases the time needed to build the model. The number of the possible views can theoretically be infinite. Therefore, it makes sense to try to reduce the number of views to a minimum while preserving a certain accuracy of the model, especially in applications for which the performance is an important issue. This paper shows an approach to Next View Planning for Shape from Silhouette for 3D shape reconstruction with minimal different views. Results of the algorithm developed are presented for both synthetic and real input images.

1 Introduction

One possibility for obtaining multiple views is to choose a fixed subset of possible views, usually with a constant step between two neighboring views, independent on the shape and the complexity of the object observed. This is illustrated in Figures 1a and 1b which show a reconstruction of a corner of a square by drawing lines from the point O with a constant angle between two lines and connecting the points, where the lines intersect the square. We can see, that the corner reconstructed using 9 lines (Figure 1b) looks "better" than the one reconstructed using 5 lines (Figure 1a), but also that neither of these two methods was able to reconstruct the corner perfectly. In addition to this, some of the views (20° in Figure 1a

[1)] 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.

and 10°, 20°, 30°, 60° and 70° in Figure 1b) could have been omitted — without them the reconstruction of the corner in Figures 1a and 1b would have been exactly the same.

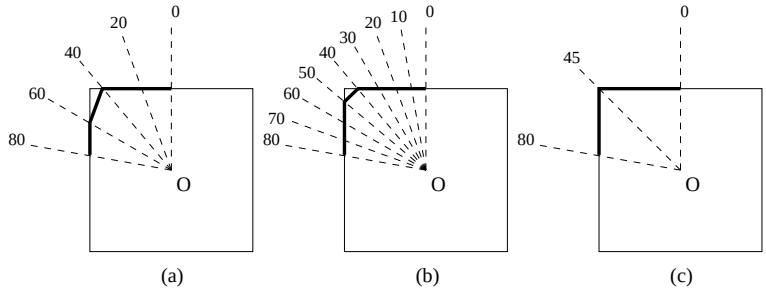

Figure 1: Reconstruction of a square corner

This simple example illustrates the need for selection of views based on the features of the object, called *Next View Planning* (*NVP*). For the square from Figure 1, if we had a way of selecting the significant views only, we could reconstruct the corner of the square perfectly using 3 views only, as shown in Figure 1c. A thorough survey of Next View Planning, also called *Sensor Planning*, is given in [16].

Our idea was to implement a simple and straight-forward NVP algorithm which will at least preserve the accuracy of models built using all possible views while reducing the number of views significantly. In most of the object reconstruction tasks which involve some kind of Next View Planning, the NVP algorithm is part of the model building process and it is guided by some features of the partial model built based on preceding views. In our 3D modeling approach the acquisition of multiple views of an object and the actual object reconstruction are separated tasks. The modeling algorithm takes the images acquired as input and does not perform any view planing itself. Therefore, our goal was to design an NVP algorithm, which does not need the partial model but uses only the features of the images acquired.

The acquisition system consists of a turntable, two cameras and one laser. The cameras and the laser are fixed while the turntable can rotate around its rotational axis. That means, our system has *one* degree of freedom. Having the constraint of using image features only, we propose a simple approach which takes only the current and the preceding image to decide what the next rotational step of the turntable will be. It defines normalized metrics for comparison of the current and the preceding image. If the change is less than or equal to the maximal allowed change then the step is doubled. If the change is higher than the maximal change, then the current image is discarded and the turntable moves back by half the current step. In special cases where doubling the step exceeds the maximum or halving the step falls below the minimum, the new step is set to the maximum or minimum, respectively.

Our approach is based on the work of Liska [10], who uses a system consisting of two lasers projecting a plane onto the viewing volume and a turntable. The next best view (the next position of the turntable) is computed based on information from the current and the preceding

scan. In each of the two scans the surface point farthest from the turntable's rotational axis is detected as well as the corresponding point in the other scan. The pair of points with the greater change in the distance from the rotational axis is used to determine whether the current turntable step should be enlarged or made smaller.

This paper is organized as follows: Section 2 describes the basic Shape from Silhouette and Shape from Structured Light method used to perform the 3D model reconstruction and Section 3 presents the Next View Planning method developed. Experimental results with both synthetic and real data are given in Section 4. At the end of the paper conclusions are drawn.

2 Acquisition Techniques

Shape from Silhouette (SfS) is a method for automatic construction of a 3D model of an object based on a sequence of images of the object taken from multiple views, where the object's silhouette represents the only interesting feature of an image [15, 13]. The object's silhouette in each view (Figure 2a) corresponds to a conic volume in 3D space (Figure 2b). A 3D model of an object (Figure 2c) can be obtained by intersecting the conic volumes which is also called *Space Carving* [9]. Multiple views of the object can be obtained either by moving the camera around the object or by moving the object inside the camera's field of view. In our approach the object rotates on a turntable in front of a stationary camera. 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.

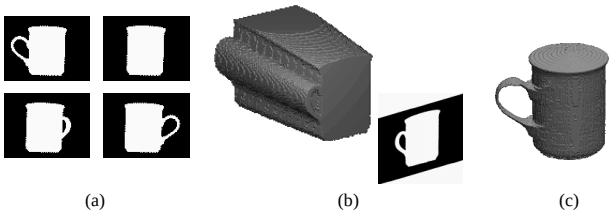

Figure 2: Image silhouettes (a), a conic volume (b) and the final model (c)

There has been much work on the construction of 3D models of objects from multiple views [1, 11, 4, 13]. Szeliski [15] 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. Wong and Cipolla [17] use uncalibrated silhouette images and recover the camera positions and orientations from circular motions. In recent years there have been also SfS approaches based on video sequences [5, 3]. The work of Szeliski [15] was used as a base for the SfS part of the method.

Shape from Structured Light (SfSL) 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 [7] or a ray or plane of laser light [10]. For every pattern an image of the scene is taken. This image, together with the knowledge about the pattern and its position relative to the camera are used to calculate the coordinates of points belonging to the surface of the object. This process is also called *active triangulation* [2, 6]. If the geometry between the laser plane and the image is known, then each 2D image point belonging to the laser line corresponds to exactly one 3D point on the surface of the object.

A strength of SfSL is that it can reconstruct any kind of concavities on the surface of the object, as long as the projected light reaches these concavities and the camera detects it. However, this method suffers from camera and light occlusions [10], resulting in incomplete surface models. The main problem 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 SfSL 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 object using our model, any intermediate surface models should be avoided, because of the problems of conversion into a volumetric model. Therefore, our approach builds a single volumetric model from ground up, using both underlying methods.

3 Next View Planning Approach

The only information provided by a pixel in a silhouette image is whether the pixel represents the object or the background. Following the notation common in NVP, we define a pixel representing the object as *seen* and a pixel representing the background as *empty*. Note that in a silhouette image there are no occlusions — the value of a pixel depends only on whether, in the conic volume defined by the pixel, there is a 3D point belonging to the object. Therefore, there can not be any *unseen* pixels, i.e., pixels for which we can not be sure whether they should be marked as seen or empty. In a binarized silhouette image all white pixels are seen and all black pixels empty. Therefore, our NVP algorithm binarizes an acquired image and compares two binary images in the following way: it counts all pixels which are seen in one and empty in the other image; in order to normalize this value, it is divided by the number of pixels which are seen in at least one of the images.

For SfSL images we follow the same idea — we mark the pixels of the current and the preceding image as *seen*, *empty* or *unseen*, and count pixels which are seen in one and empty in the other image. A SfSL input image contains a curve representing the intersection of the laser plane and the object. Our NVP algorithm compares two consecutive images by counting pixels which are seen in one and empty in the other image. This number is normalized by dividing it by the number of pixels which are seen in at least one of the two images, but not unseen in the other. In other words, because of uncertainty associated with the unseen pixels, they are completely disregarded by our NVP algorithm.

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 fit perfectly into the camera model. As synthetic object, we created a virtual cuboid with dimensions $100 \times 70 \times 60$ mm. For tests with real objects we used 6 objects: a metal cuboid, a wooden cone, a globe, a coffee cup, and two archaeological vessels. The real volume of the first 3 objects can be computed analytically, for the other objects we can only compare the bounding cuboid of the model and the object.

The user definable parameters for NVP are the maximal and the initial step between two neighboring views, as well as the maximal allowed difference between them. The parameter with the greatest impact on the number of the views selected is the difference between two images. For all objects presented the range is from 2–15%. It was low for highly symmetrical objects (the cuboids and the cone) and high when the object was not placed in the center of the turntable. For all objects the maximal step was set to 16° and the initial to 4°.

In order to evaluate the NVP-based models, we compared them with models built with a fixed number (60) of equiangular views and with models built using all 360 possible views. We expect to see that the volume of NVP-based models is closer to the volume of models built using all views than the models built with equiangular views. Figure 3 shows the models built and Table 1 summarizes the results.

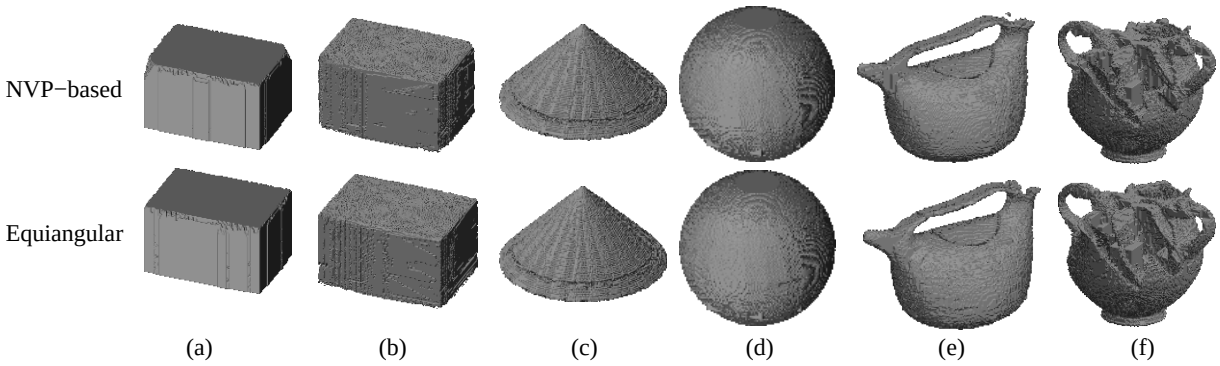

Figure 3: Comparison of models built using NVP-based and equiangular views

The results in Table 1 indicate that there is no significant difference between the volume computed using NVP-based and equiangular views for any of the objects. This can be expected for objects with asymmetric, highly detailed surfaces (the vessels), or completely rotationally symmetric objects (cone or globe). For simply shaped, but asymmetrical objects, such as the cuboids and the cup, a certain increase in the accuracy of the models built using NVP could be expected.

object	*view selection*	*#views*	*dimensions (mm)*	*volume (mm^3)*	*error*
synthetic	all	360	$100.0 \times 70.0 \times 60.0$	420 000	—
cuboid	NVP-based	54	$103.5 \times 74.0 \times 60.0$	436 666	+3.97
(Fig. 3a)	equiangular	60	$104.0 \times 73.0 \times 60.0$	434 248	+3.39
real	all	360	$101.0 \times 71.0 \times 60.0$	384 678	—
cuboid	NVP-based	54	$101.6 \times 72.3 \times 60.0$	397 937	+3.45
(Fig. 3b)	equiangular	60	$101.6 \times 71.9 \times 59.5$	397 684	+3.38
	all	360	$150.1 \times 149.4 \times 77.5$	435 180	—
cone	NVP-based	24	$151.6 \times 151.6 \times 76.5$	462 155	+6.20
(Fig. 3c)	equiangular	60	$151.6 \times 152.2 \times 76.5$	462 207	+6.21
	all	360	$149.1 \times 148.2 \times 144.6$	1 717 624	—
globe	NVP-based	24	$150.0 \times 149.1 \times 144.6$	1 733 613	+0.93
(Fig. 3d)	equiangular	60	$150.0 \times 150.0 \times 144.6$	1 732 919	+0.89
vessel	all	360	$139.2 \times 83.2 \times 92.8$	341 733	—
#1	NVP-based	52	$139.2 \times 84.0 \times 92.8$	348 699	+2.04
(Fig. 3e)	equiangular	60	$139.2 \times 83.2 \times 92.8$	346 611	+1.43
vessel	all	360	$112.9 \times 111.8 \times 86.4$	340 739	—
#2	NVP-based	55	$113.4 \times 112.8 \times 86.3$	349 918	+2.69
(Fig. 3f)	equiangular	60	$113.4 \times 112.3 \times 86.3$	348 978	+2.42
	all	360	$111.6 \times 79.0 \times 104.3$	408 344	—
cup	NVP-based	36	$112.2 \times 80.4 \times 104.3$	417 360	+2.21
(Fig. 2c)	equiangular	60	$112.2 \times 79.7 \times 104.3$	416 726	+2.05

Table 1: Comparison of silhouette models built using all views, NVP-based views and equiangular views

In order to additionally examine our NVP algorithm, in Figure 4 we illustrate the views selected for the synthetic and real cuboid, the cone and the cup. All figures show the objects from the top view, facing the x-y plane of the world coordinate system. In Figure 4 each dashed line indicates the direction the camera was viewing from, i.e., it represents the camera's optical axis. High density scanning areas should be there where the silhouette border moves fast, e.g., when the width of the silhouette changes rapidly. This happens when an object's part which is far from the rotational axis starts or ends being visible from the camera. Figure 4a illustrates the difference between two views and the dashed lines represent the optical axis of the camera. For the cuboids (Figures 4c and 4d) these parts are its corners, for the cone (Figure 4b) there are no such parts and for the cup (Figure 4e) it is its handle.

Let us analyze each of the objects from Figure 4. For the silhouette views of the cuboids (Figures 4c and 4d) the views with the highest density are 0°–60° and 180°–240°. That makes sense, because the width of the cuboid silhouettes as defined in Figure 4 is smallest for views from 30° and 210° and largest from approximately 75°, 165°, 255° and 345°. For views close to 30° and 210° the silhouette- width is determined by the two corners close to the camera. Because of being close to the camera these corners move almost orthogonally as the turntable

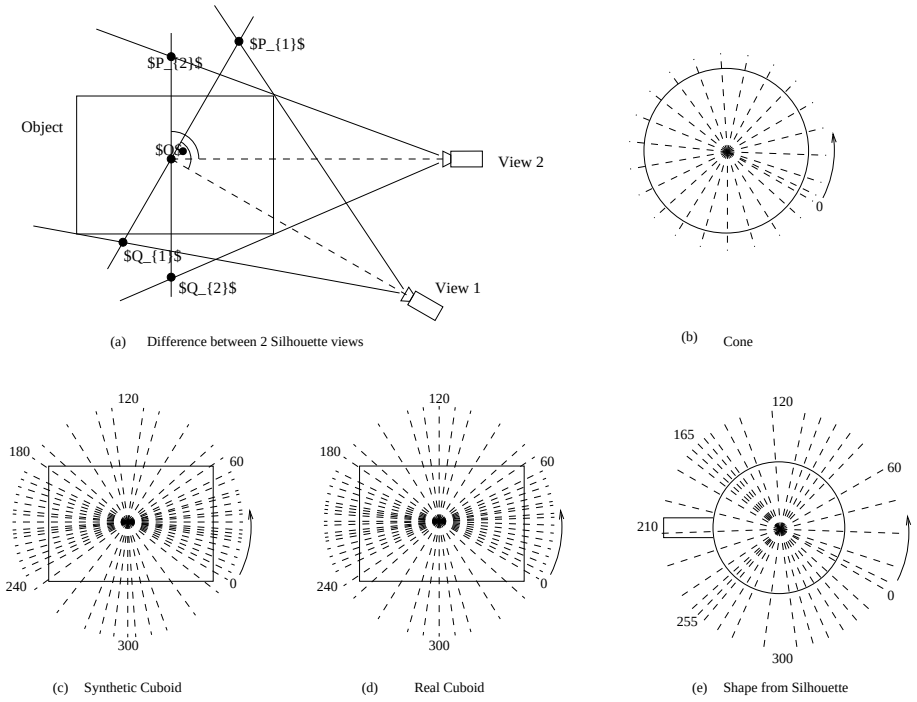

Figure 4: Analysis of selected views for cuboids, cone and cup

moves. Consequently the silhouette- width changes rapidly and the scans are most dense in these areas. For the views of the cone (Figures 4b) all views look nearly the same, so the step between two views was constantly equal to the maximal allowed step. The step was smaller only for views close to 0°, solely because of the starting angle being smaller than the maximal angle. For the silhouette-views of the cup (Figure 4e) high density views were taken from angles close to 165° and 255°. This was expected, because for those views the cup handle starts/ends being visible (i.e., not occluded by the body of the cup).

5 Conclusion

Our NVP algorithm did not fail in choosing the "right" views, and did not bring any significant differences into the results (measured in terms of the volume and the size of the objects) compared to the models built using an equivalent number of equiangular views. Therefore, the number of significant views was dramatically decreased while preserving the geometry of the object. Measuring the volume only is also not the best similarity measure since this does not necessarily describe correct geometry. For example, the NVP-based model of the cup in Figure 3 contains the complete handle, whereas the model built using equiangular views misses some parts close to the top of the handle. In conclusion we proofed that the NVP algorithm decreases the number of views to be computed (and thus save acquisition and computing time) for not highly structured objects.

References

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

[2] P. J. Besl. Active, optical range image sensors. *Machine Vision and Applications*, 1(2):127–152, 1988.

[3] A. Bottino and A. Laurentini. Non-intrusive silhouette based motion capture. In *Proceedings of 4th World Multiconference on Systemics, Cybernetics and Informatics*, pages 23–26, July 2000.

[4] 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, 1986.

[5] Q. Delamarre and O. Faugeras. 3D atriculated models and multi-view tracking with silhouettes. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, pages 716–721, 1999.

[6] 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.

[7] M. Kampel. Tiefendatenregistrierung von rotationssymmetrischen Objekten. Master's thesis, Vienna University of Technology, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group, Vienna, Austria, Februar 1999.

[8] 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.

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

[10] C. Liska. Das Adaptive Lichtschnittverfahren zur Oberflächenkonstruktion mittels Laserlicht. Master's thesis, Vienna University of Technology, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group, Vienna, Austria, April 1999.

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

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

[13] 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.

[14] R. Sablatnig, S. Tosovic, and M. Kampel. Combining shape from silhouette and shape from structured light for volume estimation of archaeological vessels. In R. Kasturi, D. Laurendeau, and C. Suen, editors, *Proc. of 16th International Conference on Pattern Recognition, Quebec City*, volume 1, pages 364–367. IEEE Computer Society, 2002.

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

[16] K. A. Tarabanis, P. K. Allen, and R. Y. Tsai. A survey of sensor planning in computer vision. *IEEE Transactions on Robotics and Automation*, 11(1):86–104, February 1995.

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