Next View Planning for a Combination of Passive and Active Acquisition Techniques*

Robert Sablatnig, Srdan Tosovic, and Martin Kampel
Pattern Recognition and Image Processing Group,
Institute for Computer Aided Automation,
Vienna University of Technology
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 a fusion of Shape from Silhouette, as an example of a passive 3D reconstruction technique, and Shape from Structured Light, as an example of an active 3D reconstruction technique in order to get 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 of 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 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.

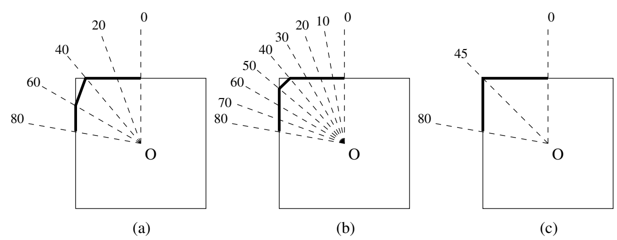


Figure 1. Reconstruction of a square corner

This simple example illustrates the need for selection of views based on the features of the object, called *Next View Planning* (*NVP*). If we had a way of selecting only the significant views for the square from Figure 1, 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 [20]. Tarabanis et al. [20], summarize the NVP problem as follows: "*Given the information about the environment (e.g., the object under observation, the available sensors) as well as the information about the task that the vision system is to accomplish (i.e., detection of certain object features, object recognition, scene reconstruction, object manipulation), develop strategies to automatically determine sensor parameter values that achieve this task with a certain degree of satisfaction*". Following this definition, in order to design an NVP algorithm for a given computer vision task, one has to identify the sensor parameters which can be manipulated (e.g., the position of the camera) and define the "degree of satisfaction", i.e., construct a metric for the evaluation of the parameter values proposed. The number of

* 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.

parameters that can be manipulated is also called the number of *degrees-of-freedom*. Increasing number of degrees of freedom increases the complexity of an NVP algorithm.

There are several computer vision tasks which can incorporate an NVP problem, differing in the necessary amount of an a priori knowledge about the object, the sensors and the environment. Maver and Bajcsy [13] proposed an NVP algorithm for an acquisition system consisting of a light stripe range scanner and a turntable. They represent the unseen portions of the viewing volume as $2\frac{1}{2}$D polygons. The polygon boundaries are used to determine the visibility of unseen portions from all candidate next views. The view which can see the largest area unseen up to that point is selected as the next best view.

Connolly [6] uses an octree to represent the viewing volume. An octree node close to the scanned surface was labeled as *seen*, a node between the sensor and this surface as *empty* and the remaining nodes as *unseen*. Next best view was chosen from a sphere surrounding the object. Connolly proposed two NVP algorithms: one called *planetarium*, which used a form of ray tracing to determine the number of unseen nodes from each candidate view and selected the one seeing the most unseen nodes, and a *normal* algorithm, which selected the next best view from 8 candidate positions only and did not take occlusions into account, and therefore was significantly faster.

Whaite and Ferrie [22] use the range data sensed so far to build a parametric approximate model of the object. The view from which the data fits the current model the worst is chosen as the next best view. This approach does not check for occlusions and does not work well with complex objects because of limitations of a parametric model.

Pito [15] uses a range scanner, which moves on a cylindrical path around the object. He partitions the viewing volume into its *seen* and *unseen* portions, and defines the surface separating the two volume portions as *void surface*. This surface is approximated by a series of small rectangular oriented *void patches*. In his *positional space (PS)* algorithm, the next best view is chosen as the position of the scanner which samples as many void patches as possible, while resampling at least a certain amount of the current model.

Similar to [15] Reed and Allen [17] propose a range scanner and a turntable system that uses an incremental modeler and a sensor planer that work in an interleaved fashion. A "rough" model is acquired first, the occlusion boundary defines the *unseen* surface, which serves as the criterion for the next best view. The planning task tries to maximize the surface area of the occlusion boundary that is imaged in each sensing operation.

Our idea was to implement a simple and straight-forward NVP algorithm, which can perform locally so that the turntable movements are minimal and works with both data sources, active and passive triangulation data. Furthermore, the method should 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. As in [17] in our 3D modeling approach the acquisition of multiple views of an object and the actual object reconstruction are separated tasks, since we have two different models to be combined into one and two different viewpoints (active and passive system). 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 and is simple and fast.

The acquisition system consists of a turntable with a diameter of 50 cm, whose desired position can be specified with an accuracy of $0.05°$ (Figure 2a); two monochrome CCD-cameras (*Camera-1* in Figure 2 is used for acquiring the images of the object's silhouettes and *Camera-2* in Figure 2 for the acquisition of the images of the laser light projected onto the object); a backlighting system (Figure 2e) used to illuminate the scene for the acquisition of the silhouette of the object; and one prism equipped laser (Figure 2d) used to project a light plane onto the object. 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.

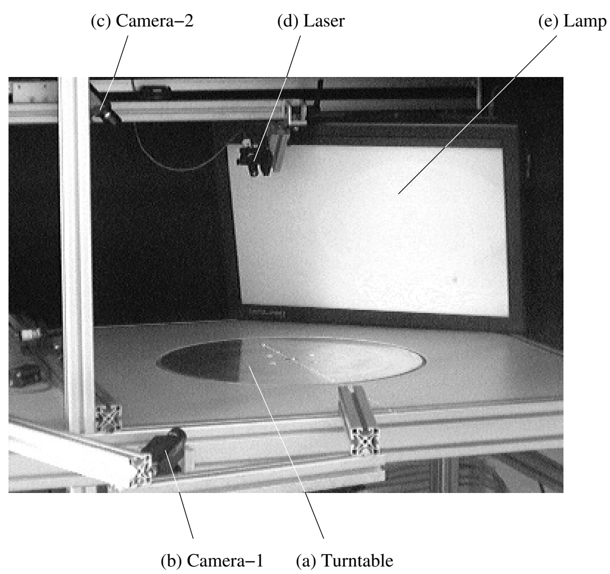


Figure 2. Acquisition System

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.

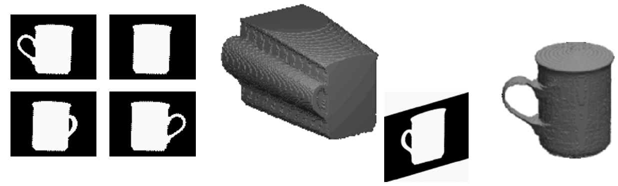
Our approach is based on the work of Liska [11], who uses a system consisting of two lasers projecting a plane onto the viewing volume and a turntable. The next best view (the next position of the turntable) is computed based on information from the current and the preceding scan. In each of the two 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 minimized.

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 for both reconstruction methods. 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 Techniques

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, where the object's silhouette represents the only interesting feature of an image [19, 16]. The object's silhouette in each view (Figure 3a) corresponds to a conic volume in 3D space (Figure 3b). A 3D model of an object (Figure 3c) can be obtained by intersecting the conic volumes, which is also called *Space Carving* [10]. 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. Shape from Silhouette 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.

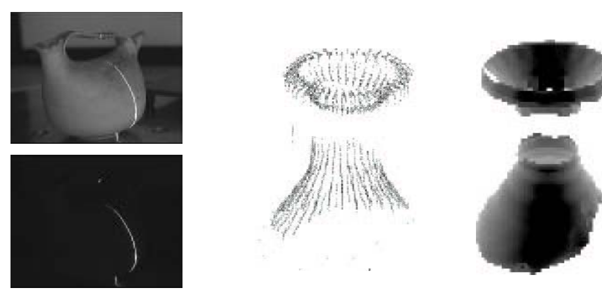
There has been much work on the construction of 3D models of objects from multiple views [1, 12, 5, 16]. Szeliski [19] first creates a low resolution octree model quickly and then refines this model iteratively, by intersecting each new silhouette with the already existing model. Niem [14] uses pillar-like volume elements instead of an octree for the model representation. Wong and Cipolla [23] use uncalibrated silhouette images and recover the camera positions and orientations from circular motions. In recent years there have been also Shape from Silhouette approaches based on video sequences [7, 3]. The work of Szeliski [19] was used as a base for the Shape from Silhouette part of the method (for details see [21]).



(a) (b) (c)

Figure 3. Image silhouettes (a), a conic volume (b) and the final model (c)

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 [9] or a ray or plane of laser light [11]. 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, 8].



(a) (b) (c)

Figure 4. Projection of laser plane (a), cloud of points (b) and reconstructed surface (c)

The Shape from Structured Light method used in our approach is based on the projection of laser planes onto the object (Figure 4a). 3D points obtained through active triangulation using all views represent a cloud of points belonging to the object's surface (Figure 4b). This cloud of points can be used to create a smooth surface representation (Figure 4c). A strength of Shape from Structured Light is that it can reconstruct any kind of concavities on the surface of the object (see the top of the amphora in Figures 4b and 4c), as long as the projected light reaches these concavities and the camera detects it. However, this method often suffers from camera and light occlusions [11], resulting in incomplete surface models.

The combination of the Shape from Silhouette method with the Shape from Structured Light method was performed in order to provide volumetric data of objects with concavities and to eliminate the drawbacks of each of the methods since they complement one another (see [18] for details). 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 Shape from Structured Light surface model and a Shape from Silhouette 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 to a volumetric model. Therefore, our approach proposes building a single volumetric model up from the ground, using both underlying methods (for details see [18]). For the volumetric model we use an octree [4], which is a tree-formed data structure used to represent 3D objects. Each node of an octree represents a cube subset of a 3D volume. The octree contains binary information in the leaf nodes and is therefore 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. The octree representation has several advantages [4]: 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. Since both acquisition methods use the same modeling scheme, our NVP method can be applied to both techniques independently at the same time.

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. This condition holds, if the angle between two views is not too large, since it is a local property.

With this metric definition, if two silhouette images are identical, the change is 0, and if the silhouettes do not intersect at all, it is 1. Note that calculating the change uses features of the images only and none of the information about the geometry of the acquisition system. This means that the system does not need to be calibrated prior to applying the NVP algorithm. Our NVP approach performs these steps:

1. Parameters are initialized. The user sets the initial step α_{init} and the maximal step α_{max} ($\alpha_{init} \leq \alpha_{max}$), as well as the maximal allowed change C_{max} between two subsequent images. This change is assumed to be normalized, i.e., $0 \leq C_{max} \leq 1$. The minimal step α_{min} is implied by the resolution of the turntable (1° for our turntable).

2. The first image I_1 is taken. The current step α_{curr} is set to the initial value: $\alpha_{curr} = \alpha_{init}$. Number of acquired views n is set to one: $n = 1$.

3. If the turntable already has made a complete revolution of 360°, we are done. Otherwise, the turntable is rotated by the angle α_{curr}, the image I_{n+1} is taken and we continue with Step 4.

4. The change C_{curr} between the images I_{n+1} and I_n is evaluated. If $C_{curr} \leq C_{max}$ or $\alpha_{curr} = \alpha_{min}$ the image I_{n+1} is accepted, jump to Step 6. Otherwise the image I_{n+1} is discarded, continue with Step 5.

5. The step α_{curr} is halved: $\alpha_{curr} = \frac{1}{2} \cdot \alpha_{curr}$. If α_{curr} becomes smaller than α_{min} it is set to α_{min}. The turntable is rotated by $-\alpha_{curr}$ (i.e., back by the half of the previous step). Go back to Step 4.

6. Increment the image counter n by one and double the step α_{curr}: $n = n + 1$, $\alpha_{curr} = 2 \cdot \alpha_{curr}$. Jump back to Step 3.

For Shape from Structured Light 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 Shape from Structured Light input image contains a curve representing the intersection of the laser plane and the object. How do we decide which pixels are seen/empty/unseen? If we denote the source point of the laser with P, the image pixels produced by the laser line with R_x and all possible object points Q_i (Figure 5), and draw a line from P going through R_x, we can differentiate between three types of points:

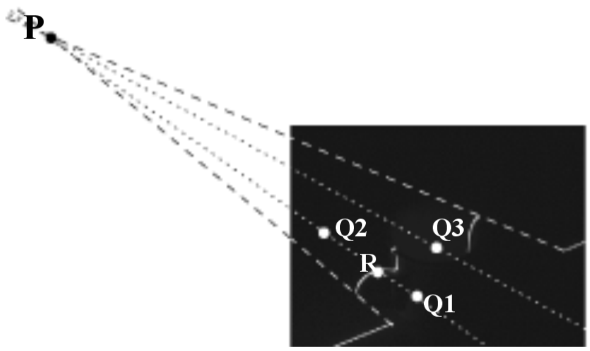


Figure 5. Seen, empty and unseen pixels in laser images

If the line intersects the laser curve before it reaches R_x (or exactly at R_x), then Q_i is *below* the surface of the object and we mark it as seen (point Q_1 in Figure 5); if the line intersects the laser curve after the point R_x, then Q_i is *above* the object's surface and we mark it as empty (Q_2 in Figure 5); finally, if the line does not intersect the laser curve at all, then Q_i is occluded by a part of the object outside the current laser plane and we can not say whether it is above or below the surface, so we mark it as unseen (Q_3 in Figure 5). 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 such that the 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

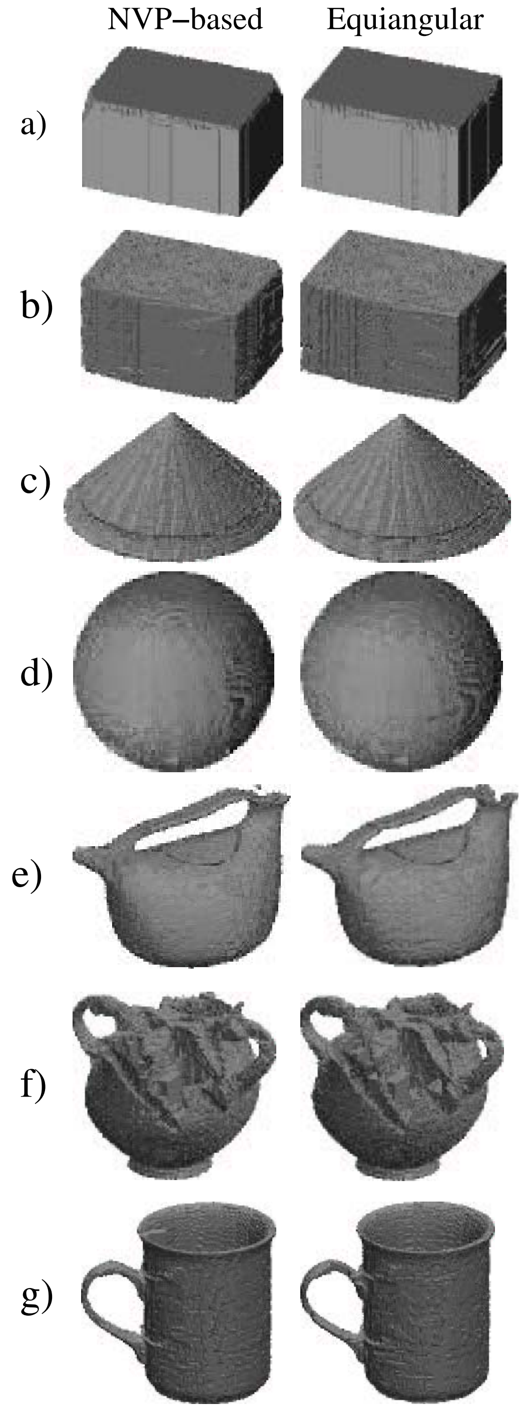


Figure 6. Comparison of models built using NVP-based and equiangular views

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 $4°$.

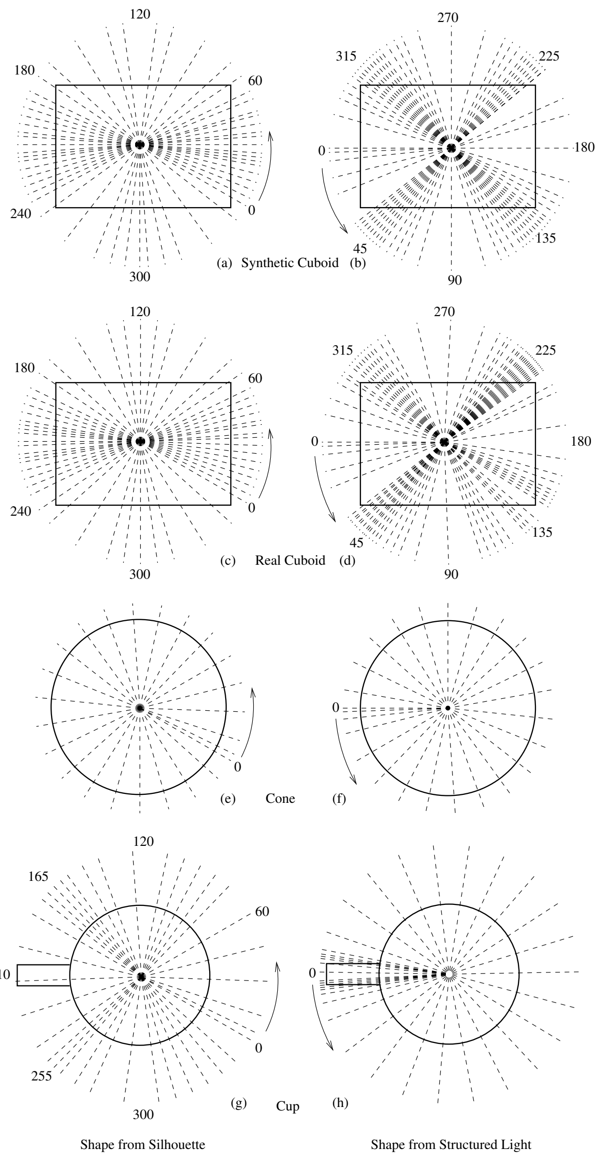


Figure 7. Analysis of selected views for cuboids, cone and cup

In order to evaluate the NVP-based models, we compare them with models built with a fixed number (60) of equiangular views and with models built using all 360 possible views. For the laser method we took also different number of equiangular views since for some objects fewer views (like the cone) or more views (like vessel 2 which is highly structured) are necessary to obtain comparable results. 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 6 shows the models built. and Table 1 summarizes the results.

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, such as the vessels or completely rotationally symmetric objects, such as the cone or the 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. Certainly, for highly structured objects like vessel 2, the method produced even more views than in the equiangular case to produce an ideal number of scans. This can be expected since the simple NVP method scans almost all highly structured parts at the lowest step width.

In order to additionally examine our NVP algorithm, in Figure 7 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.

For Shape from Silhouette views (Figures 7a, 7c, 7e and 7g) each dashed line indicates the camera viewing direction, i.e., it represents the camera's optical axis. High density scanning areas should be those for which 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. For the cuboids (Figures 7a and 7c) such parts are its corners, for the cone (Figure 7e) there are no such parts and for the cup (Figure 7g) it is its handle.

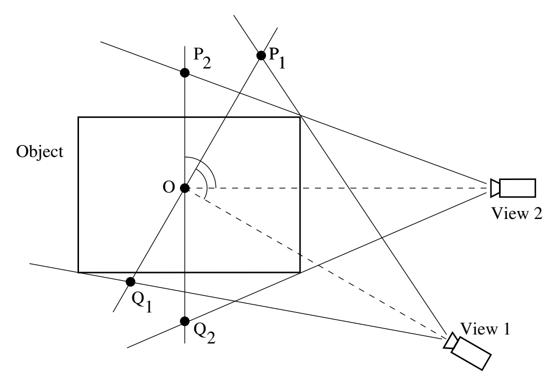


Figure 8. Difference between two silhouette views

For the purpose of better understanding of the selected Shape from Silhouette views in Figure 7, Figure 8 illustrates the difference between two views, where each dashed line represents the optical axis of the camera. If we define O from Figure 8 as the point representing the rotational axis of the turntable, then we can view the lines $\overline{P_1Q_1}$ and $\overline{P_2Q_2}$ as the width of the silhouette in views 1 and 2, respectively. The parts of an object for which this width changes signifi-

Proceedings of the Fourth International Conference on 3-D Digital Imaging and Modeling (3DIM'03)
0-7695-1991-1/03 $ 17.00 © 2003 IEEE

object	*view selection*	*#silh.views*	*#las.views*	*dimensions (mm)*	*volume* (mm^3)	*error*
synthetic	all	360	360	$100.0 \times 70.0 \times 60.0$	420 000	—
cuboid	NVP-based	54	71	$103.5 \times 74.0 \times 60.0$	436 666	+3.97%
(Fig. 6a)	equiangular	60	90	$104.0 \times 73.0 \times 60.0$	434 248	+3.39%
real	all	360	360	$101.0 \times 71.0 \times 60.0$	384 678	—
cuboid	NVP-based	54	79	$101.6 \times 72.3 \times 60.0$	397 937	+3.45%
(Fig. 6b)	equiangular	60	90	$101.6 \times 71.9 \times 59.5$	397 684	+3.38%
	all	360	360	$150.1 \times 149.4 \times 77.5$	435 180	—
cone	NVP-based	24	25	$151.6 \times 151.6 \times 76.5$	462 155	+6.20%
(Fig. 6c)	equiangular	60	60	$151.6 \times 152.2 \times 76.5$	462 207	+6.21%
	all	360	360	$149.1 \times 148.2 \times 144.6$	1 717 624	—
globe	NVP-based	24	25	$150.0 \times 149.1 \times 144.6$	1 733 613	+0.93%
(Fig. 6d)	equiangular	60	60	$150.0 \times 150.0 \times 144.6$	1 732 919	+0.89%
vessel	all	360	360	$139.2 \times 83.2 \times 92.8$	341 733	—
#1	NVP-based	52	45	$139.2 \times 84.0 \times 92.8$	348 699	+2.04%
(Fig. 6e)	equiangular	60	60	$139.2 \times 83.2 \times 92.8$	346 611	+1.43%
vessel	all	360	360	$112.9 \times 111.8 \times 86.4$	340 739	—
#2	NVP-based	55	148	$113.4 \times 112.8 \times 86.3$	349 918	+2.69%
(Fig. 6f)	equiangular	60	120	$113.4 \times 112.3 \times 86.3$	348 978	+2.42%
	all	360	360	$111.6 \times 79.0 \times 104.3$	408 344	—
cup	NVP-based	36	34	$112.2 \times 80.4 \times 104.3$	417 360	+2.21%
(Fig. 6g)	equiangular	60	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

cantly from one view to the next need to be scanned with a higher density.

Shape from Structured Light views (Figures 7b, 7d, 7f and 7h) are easier to understand. A dashed line in these figures represents the laser plane projected onto the object for that view. High density scanning areas should be those where the distance between the rotational axis and the intersection point of the laser plane and the object surface changes rapidly. For the cuboids (Figures 7b and 7d) such areas are the ones around the cuboids' corners, for the cone (Figure 7f) these areas do not exist, and for the cup (Figure 7h) they lie around the cup handle.

Let us analyze each of the objects from Figure 7. For the silhouette views of the cuboids (Figures 7a and 7c) 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 8 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 (see View 2 in Figure 8). Because of being close to the camera these corners move almost orthogonally as the turntable moves, so the silhouette width changes rapidly here and the scans are most dense in these areas. The laser views of the cuboids (Figure 7b and 7d) are more dense close to the corners, as expected, but it can also be seen that for both synthetic and the real cuboid several corners were missed, because the step was too large, but the NVP algorithm did not see it — for example, the next left and next right view of the lower right corner intersect the cuboid surface at almost the same distance from the rotational center.

For both silhouette and laser views of the cone (Figures 7e and 7f) 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 7g) high density view were taken from angles close to $165°$ and $255°$. This is expected, because for those views the cup handle starts/ends being visible (i.e., not occluded by the body of the cup). The laser views (Figure 7h) are dense only in areas where the projected laser plane "jumps" from the body of the cup to its handle and back, just as expected.

5 Conclusion and Outlook

Obviously, our NVP algorithm did not fail in choosing the "right" views (except for the laser views of corners of the cuboids), and did not bring any significant differences in 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 for not structured objects. For highly structured objects the method produces more views, since the structure of the surface should be reconstructed best. Measuring the volume only is not the best similarity measure too, since this does not necessarily describe correct geometry. For example, the NVP-based model of the cup in Figure 6 contains the complete handle, whereas the model built using equiangular views misses some parts close to the top of the handle. In conclusion we have shown that the NVP algorithm for Shape from Silhouette and Shape from Structured Light is able to decrease the number of views to be computed (and thus save acquisition and computing time) for non highly structured objects or objects that are highly structured only at specific regions (like a cup with handle). As a consequence we want to test our NVP algorithm with complex, asymmetric synthetic objects and would like to extend the method to two degrees of freedom.

References

[1] H. Baker. Three-dimensional modelling. In *Proc of 5th Intl. Conf. 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] 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.

[5] C. H. Chien and J. K. Aggarwal. Volume/surface octrees for the representation of three-dimensional objects. *CVGIP*, 36:100–113, 1986.

[6] C. Connolly. The determination of next best views. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 432–435, 1985.

[7] Q. Delamarre and O. Faugeras. 3D atriculated models and multi-view tracking with silhouettes. In *Proc. of 7th ICCV*, pages 716–721, 1999.

[8] 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.

[9] M. Kampel. Tiefendatenregistrierung von rotationssymmetrischen Objekten. Master's thesis, Vienna University of Technology, Institute of Computer Aided Automation, 1999.

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

[11] C. Liska. Das Adaptive Lichtschnittverfahren zur Oberflächenkonstruktion mittels Laserlicht. Master's thesis, Vienna University of Technology, Institute of Computer Aided Automation, 1999.

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

[13] J. Maver and R. Bajcsy. Occlusions as a guide for planning the next view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5):417–432, May 1993.

[14] 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.

[15] R. Pito. A solution to the next best view problem for automated surface acquisition. *IEEE-PAMI*, 21(10):1016–1030, October 1999.

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

[17] M.K. Reed and P.K. Allen. Constraint-based sensor planning for scene modeling. *IEEE-PAMI*, 22(12):1460–1467, 2000.

[18] 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 ICPR*, volume 1, pages 364–367, 2002.

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

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

[21] S. Tosovic and R. Sablatnig. 3D modeling of archaeological vessels using shape from silhouette. In *Proc. of 3rd Intl. Conf. on 3-D Digital Imaging and Modeling*, pages 51–58, May 2001.

[22] P. Whaite and F. P. Ferrie. Autonomous exploration: Driven by uncertainty. In *Proc. of CVPR*, pages 339–346, 1994.

[23] K. Y. K. Wong and R. Cipolla. Structure and motion from silhouettes. In *Proc. of 8th ICCV*, pages 217–222, 2001.