Technical Report

Pattern Recognition and Image Processing Group
Institute of Computer Aided Automation
Vienna University of Technology
Favoritenstr. 9/1832
A-1040 Vienna AUSTRIA
Phone: +43 (1) 58801-18351
Fax: +43 (1) 58801-18392
E-mail: tos@prip.tuwien.ac.at
URL: http://www.prip.tuwien.ac.at/

PRIP-TR-066 July 10, 2001

Construction of 3D Models of Objects Using Combination of Shape from Silhouette and Shape from Structured Light

Specification of the master thesis

Srdan Tosovic

Abstract

This work proposes a method of three-dimensional reconstruction of objects using a combination of two existing methods, *Shape from Silhouette* and *Shape from Structured Light*. The purpose of combining these two methods is to overcome the disadvantages of one algorithm by applying the other and construct a 3D model of an object which will be more accurate and precise than a model acquired by applying one of the algorithms only.

Contents

List of Figures

1 Introduction

The 3D acquisition method proposed in this thesis is being done as part of the project *Computer Aided Classification of Ceramics* [MS96, KS99], which is performed by the Pattern Recognition and Image Processing Group at the Institute of Computer Aided Automation at the Vienna University of Technology in cooperation with the Institute of Classical Archeology at the University of Vienna, with the goal of providing an objective and automated method for classification and reconstruction of archaeological pottery.

Three-dimensional reconstruction of archaeological vessels is interesting for archaeologists for several reasons. The surface of a 3D model of a vessel can be unwrapped easily and can help archaeologists illustrate its decoration, which is important for classification. Furthermore, the volume of a vessel can be estimated by calculated the volume of its 3D model, which also allows more precise classification [OTV93].

The method proposed is a combination of two existing methods which have already been implemented at PRIP, *Shape from Silhouette* [Tos00] and *Shape from Structured Light* [Lis99]. Both of these methods have their strengths and weaknesses, as discussed below. The purpose of combining the two methods is to overcome the weaknesses of one method by applying the other so that a 3D model of an object can be constructed which is more accurate and precise than a model acquired by applying only one of the methods.

The remainder of this section describes the two underlying methods, *Shape from Silhouette* and *Shape from Structured Light*. Section 2 describes the acquisition system which will be used for the implementation. Section 3 defines the goals of this thesis and gives an estimation for its time schedule, followed by conclusions in Section 4.

1.1 Shape from Silhouette

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 an image [Sze93, Pot87]. The object's silhouette in each input image corresponds to a conic volume in the object real-world space (Figure 1). A 3D model of the object can be built by intersecting the conic volumes from all views.

Shape from Silhouette is a computationally simple algorithm — it employs only basic matrix operations for all transformations — and it requires only a camera as equipment, so it can be used to obtain a quick initial model of an object which can then be refined by other methods. It 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. It can also be used to estimate the volume of an object.

Multiple views of the object can be obtained either by moving the camera around the object or by moving the object within the camera's field of view. In our approach the

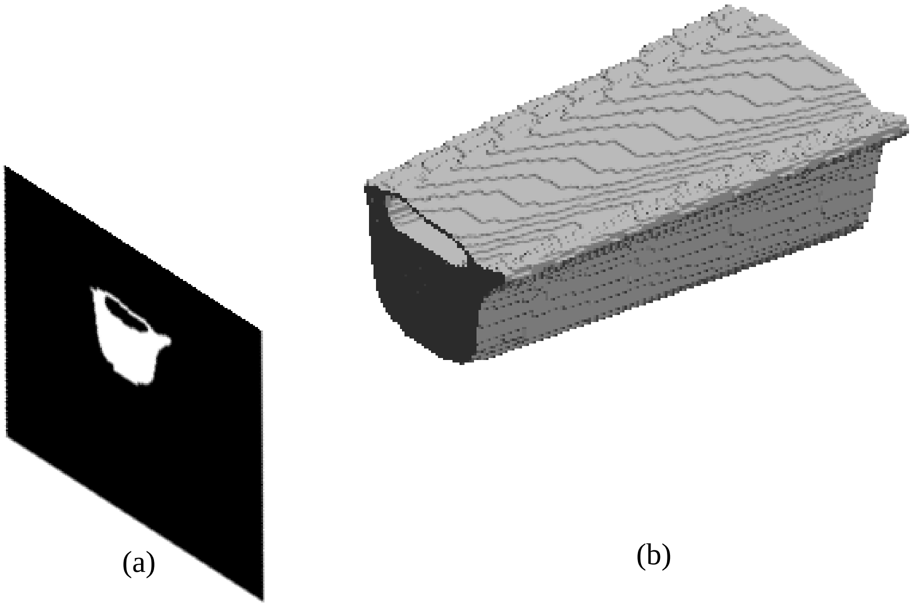


Figure 1: Image silhouette (a) and the corresponding conic volume (b)

object rotates on a turntable in front of a stationary camera (see Section 2).

The strength of *Shape from Silhouette*, beside its simplicity, is that it can reconstruct objects with shapes which can hardly be reconstructed by other 3D acquisition methods. Examples of such shapes are torus-like objects, as well as objects with a handle, like the vessel in Figure 2a, reconstructed by *Shape from Silhouette*, shown in Figure 2b.

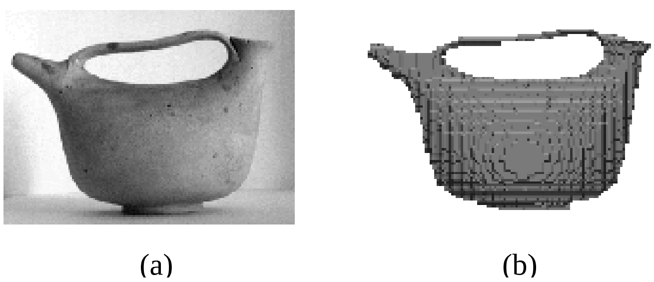


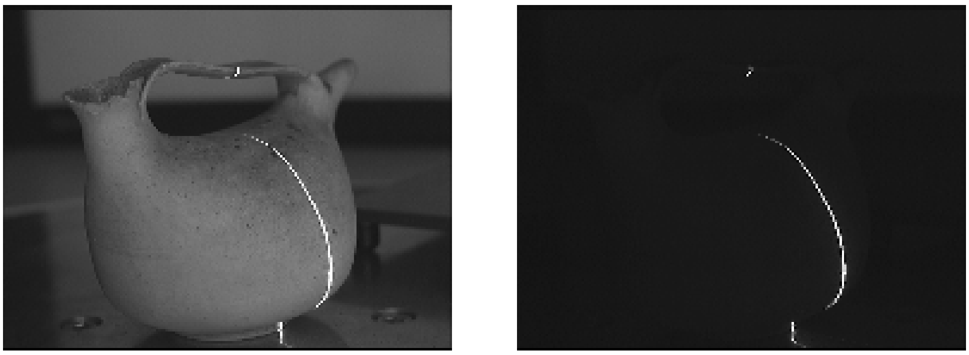
Figure 2: Archaeological vessel (a) and its 3D model (b)

The drawback of any *Shape from Silhouette* algorithm is that certain types of concavities of an object can not be reconstructed independently from the number of input images and the location from which they were taken. Those are concavities which can not be seen by any input image, such as the inside of a bowl or a cup.

1.2 Shape from Structured Light

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. For every pattern an image of the scene is taken. This image, together with the knowledge about the pattern and its relative position to the camera are used to calculate the coordinates of points belonging to the surface of the object.

There are several kinds of *Shape from Structured Light* [Joh93, YCZB98]. The one which will be used for the method proposed in this work projects a plane onto the object using laser light (Figure 3a). The image of such a scene contains only the line which represents the intersection of the object and the laser plane (Figure 3b). In order to reconstruct the whole object the laser plane has to be projected onto different parts of the object, which can be achieved by either moving the laser or moving the object. Just like for the *Shape from Silhouette* part of the method proposed, multiple views of the object are obtained by rotating the object on the turntable. A detailed description of the complete acquisition system is given in Section 2.



(a) (b)

Figure 3: Projection of laser light onto an object

A strength of this variant of *Shape from Structured Light* (using laser plane projection) is that it can reconstruct an object's concavities like the inside of a bowl, depending on the relative position of the laser and the camera to the object.

On the other hand, the concavities like the handle of a cup can only be poorly reconstructed, because it is very likely that the handle would occlude other parts of the object, so no laser projection would reach the occluded parts. This problem can partly be addressed by using multiple lasers.

2 Acquisition System

This section gives an overview of the acquisition system which will be used for the implementation of the 3D acquisition method proposed in this work. Section 2.1 describes the devices of the acquisition system and their geometrical setup. Section 2.2 defines the relevant coordinate systems and discusses which parameters need to be calculated by the calibration of the system.

2.1 Description

The acquisition system consists of the following devices:

- a turntable with a diameter of 50 *cm*, whose desired position can be specified with an accuracy of 0.05°. The turntable is used to obtain multiple views of the object observed.
- two monochrome CCD-cameras with a focal length of 16 *mm* and a resolution of 768x576 pixels. One camera is used for acquiring the images of the object's silhouettes and the other for the acquisition of the images of the laser light projected onto the object.
- two lasers which are used to project a light plane onto the object. Each laser is equipped with a prism in order to span a plane out of the laser beam. Both lasers project red light.
- one or more lamps, to illuminate the scene for the acquisition of the silhouette of the object. The object should be clearly distinguishable from the background. This can be best achieved with backlighting [HS91].

The most probable geometrical setup of the acquisition devices is shown in the Figure 4. 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 (*Camera-1* in Figure 4) lies nearly in the rotational plane of the turntable, orthogonal to the rotational axis. The camera for acquiring the projection of the laser plane (*Camera-2* in Figure 4) onto the object views the turntable from an angle of about 45° (β in Figure 4).

The two lasers are directed in such way that their laser planes completely overlap and contain the rotational axis. One laser could be used instead of one, too, but using two lasers and placing the camera between them greatly enlarges the effective acquisition area by reducing light occlusions (see Figure 5). *Camera-2* in Figure 4 views the laser plane from an angle of about 45° (α in Figure 4).

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 of one another.

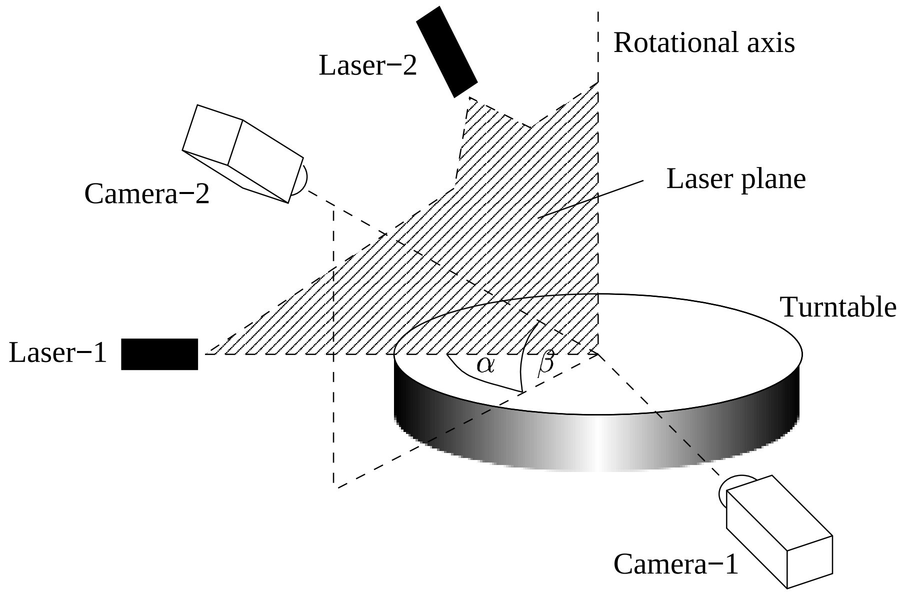


Figure 4: Geometrical setup of acquisition system

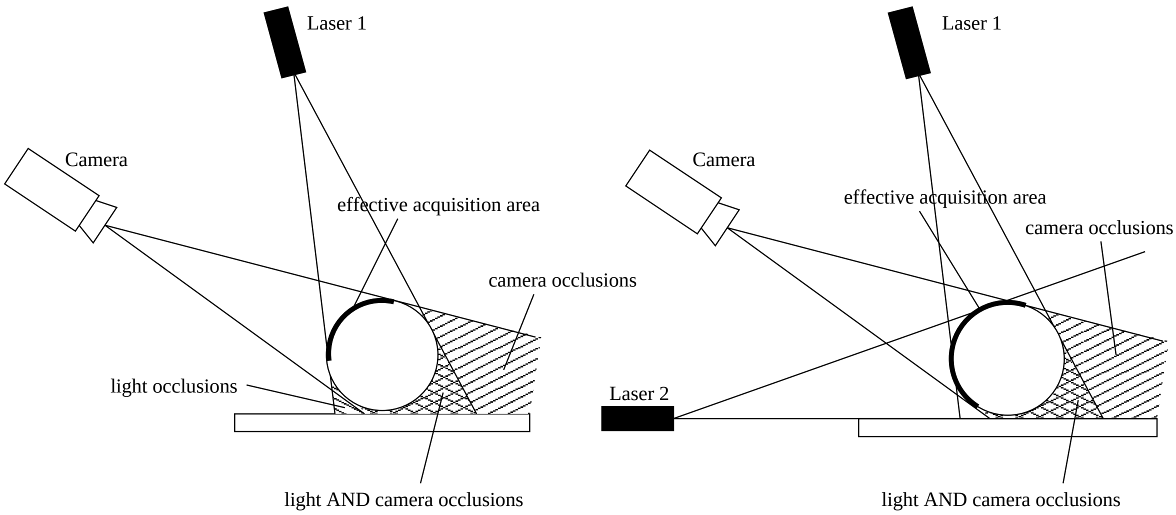


Figure 5: Light and camera occlusions

2.2 Calibration

Before 3D reconstruction of any object can take place, the acquisition system has to be calibrated, i.e. the relative positions of the cameras, the lasers and the turntable have to be computed.

For the acquisition system described in this section we define the following coordinate systems (Figure 6), all of which are right-handed:

- *object* or *real-world* coordinate system: it is rooted at the intersection of the rotational axis of the turntable and its rotational plane. Its z axis is identical to the rotational axis of the turntable. x and y axis lie in the turntable's rotational plane and their exact position depends on the camera calibration method used.

- *camera-1* coordinate system: rooted at the center of the lens of the *Camera-1* from Figure 4, with z axis being identical to the camera's optical axis, and x and y axes as shown in Figure 6.
- *camera-2* coordinate system: *Camera-2* equivalent of the camera-1 coordinate system.
- *image-1* coordinate system: 2D image coordinate system for images taken with *Camera-1*.
- *image-2* coordinate system: 2D image coordinate system for images taken with *Camera-2*.

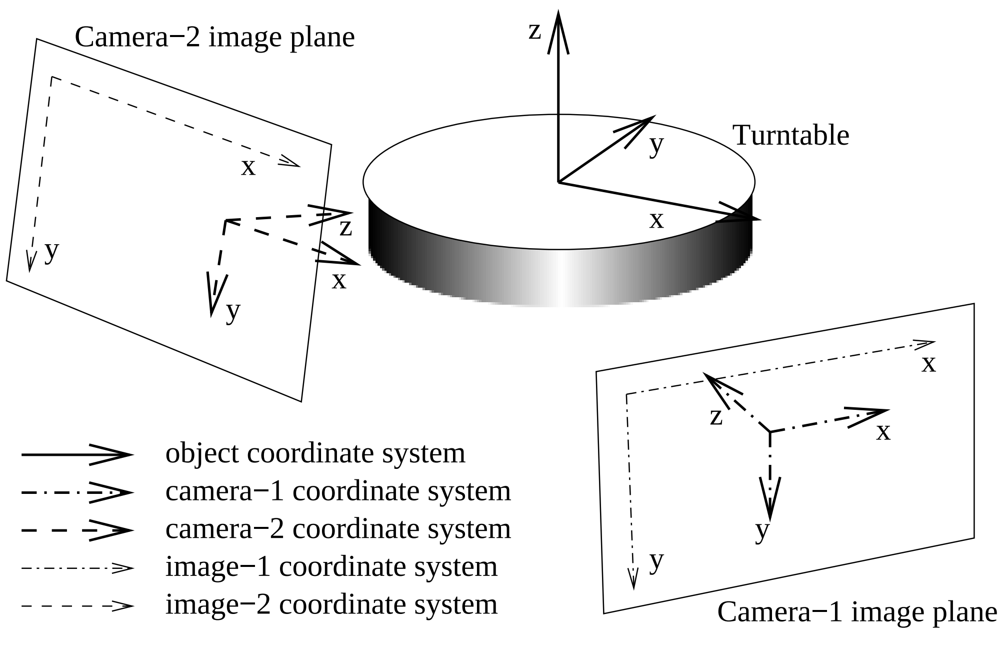


Figure 6: Relevant coordinate systems

With this definition of the relevant coordinate systems the calibration has to provide us with the following information:

1. the transformation between object and camera-1 coordinate system.
2. the transformation between object and camera-2 coordinate system.
3. the transformation between camera-1 and image-1 coordinate system.
4. the transformation between camera-2 and image-2 coordinate system.

5. the equation of the laser plane in object coordinate system.

For the calibration of the system described in this section we plan to use some variant of Tsai [Tsa86] or DLT [AAK71] calibration, whose implementations already exist at the PRIP-Group.

3 Goals and Time Schedule

The primary goal of this work is to develop and implement the new 3D acquisition method, a combination of *Shape from Silhouette* and *Shape from Structured Light*. Section 3.1 discusses the key problems and issues that need to be addressed in reaching this goal. Section 3.2 describes several enhancements that can be applied to the current implementations of the two underlying methods. These enhancements can be viewed as the secondary goal of the thesis. Finally, Section 3.3 gives a date estimate for the completion of single steps.

3.1 Combining the Two Methods

The problem of combining of *Shape from Silhouette* and *Shape from Structured Light* can be approached in several ways:

- run each of the methods separately, which provides us two independent models of the object. The final model could be constructed as an intersection of the two models.
- run one method on top of the other. For example, we could build an initial model using *Shape from Silhouette* and then refine this model using *Shape from Structured Light*.
- create a hybrid-method, i.e. build a single model of the object using both underlying methods in each step.

The first approach, while sounding intuitive, includes several problems. The main difference between *Shape from Silhouette* and *Shape from Structured Light* is that the first method builds up a volumetric model of the object and the latter one just gives a cloud of points belonging to the surface of the object. In order to intersect these two models they have to be adapted to one another. If we convert *Shape from Silhouette*'s volumetric model into some kind of surface representation, it could include some information loss and it is also not a simple task to intersect two surfaces, i.e. decide which one is "better" on which part of the object. Intersection of two volumetric models is easier, because we would only have to determine for every object voxel of one model whether it also belongs to the object in the other model, and if it does not, then delete the voxel from the model. But conversion of *Shape from Structured Light*'s cloud of surface points into a volumetric model is also a challenging task, especially because, in general, many parts of the object stay occluded during the whole acquisition process and the question is what to do with the parts of acquisition space the algorithm did not give any information about.

The second approach is similar to the first one, with the difference that we do not have to build a complete model with the method used for the refinement. However, we face the same problems of adapting the method's model representations to one another.

The hybrid-method could avoid the problem of converting one model representation to the other. One could "carefully" build a data structure representing the object by taking into account images acquired both for *Shape from Silhouette* and *Shape from Structured Light*. The current implementation of *Shape from Silhouette* builds an octree [CH88] model of the object. The option of using octree model in *Shape from Structured Light* will be investigated.

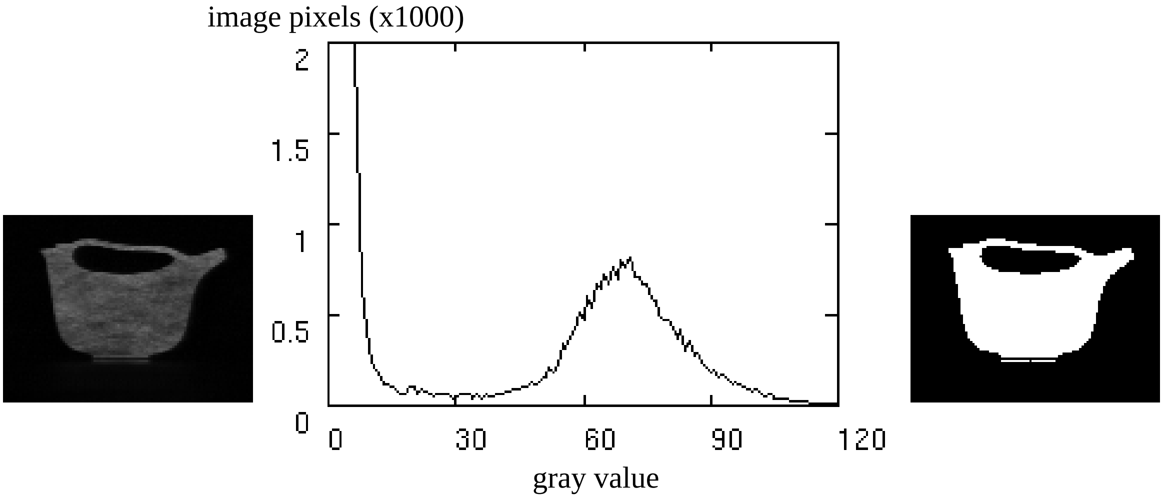
3.2 Enhancements

The current implementation of *Shape from Silhouette* [Tos00] has several weaknesses:

- the algorithm assumes that the camera lies in the rotational plane of the turntable, with its optical axis orthogonal to the turntable's rotational axis (*Camera-1* in Figure 4), making the algorithm unusable with any other position of the camera relative to the turntable.
- in order to extract the object's silhouette from input images, the user has to set the threshold manually, which, if set wrong, has a great impact on the accuracy of the output 3D model.
- another problem with the silhouette extraction is that the same threshold value is used for all input images. This problem is less significant because usually all input images are taken under similar conditions and the "perfect" threshold value does not differ a lot from one image to another.
- the angle between two neighboring views is also set manually by the user and constant throughout the image acquisition.

The assumption about the relative positions of the camera and the turntable was made by the calibration algorithm [Tos99] which was developed for the current implementation of *Shape from Silhouette*. The problem of *Shape from Silhouette* is that it uses the knowledge about the calibration for calculating projection of octree nodes into the image plane. In the approach proposed in this work the relative positions of the camera and the turntable will remain the same (Figure 4), but the approach should be independent on the calibration method used, so it can be tested and used with different positions of the camera.

The problem of the manual threshold setting for silhouette extraction from input images has already been addressed at the time of writing of this report. The threshold is now set automatically and individually for every input image. Figure 7 illustrates how the threshold is set. Based on a certain criteria, it finds a point in the histogram of the input image between the two highest histogram peaks. An input image is obtained by calculating the absolute difference between the image of the object acquired by the



(a) Input image

(b) Histogram

(c) Binary image

Figure 7: Automatic silhouette extraction

camera and the image of the object space without any objects in it. For the image in Figure 7a the calculated threshold is 32. Figure 7b shows only the interesting region of the histogram. The highest peak of the histogram is for the gray value of 0 (black) and its value is more than 17 000.

Another option for extracting object silhouettes from input images would be to use edge detection instead of thresholding. This approach could be more accurate, even a sub-pixel precision could be reached, but it is also more complex. One of the questions that would have to be answered, given the edges of the object, is how to decide what is inside and what outside the object. This question might be easy to answer for simple objects, but for more complex objects (e.g. archaeological vessels with a thin handle) it could be rather complicated. Another problem would be computer resources. In the experiments with the current implementation of *Shape from Silhouette* described in [Tos00] 1-pixel resolution of the object silhouettes was sufficient to build octree models with a depth of 8, i.e. resolution of 256^3. Creating the final output model temporarily required 16 MB of memory. If the octree depth of 9 had been used, this memory requirement would rise to 128 MB and for the depth of 10 to 1 GB. Apart from buying more memory, this problem can be addressed by optimizing the model-building algorithm. In our opinion, it still does not justify the development of a complex silhouette extraction algorithm. In the approach proposed in this work, one of the ideas is to place the camera at least 50% closer to the object than in experiments described in [Tos00], so even the 1-pixel precision of thresholding will be sufficient to build more detailed models of objects than the ones presented in [Tos00]. However, some optimizations will be done to the *Shape from Silhouette* algorithm and after the first experimental results using thresholding for silhouette extraction are available, the need for a better silhouette extraction algorithm will be reevaluated.

The problem of a constant angle between two neighboring views can be addressed by introducing of some kind of *Next View Planning* [MB93] for the calculation of the next view. This can make the number of images needed for building an accurate 3D model smaller, making the angle between two neighboring views small only for the object parts with many details. However, it can be argued that the little details usually mean little cavities on the object which can not be seen in the silhouette of the object independent of the viewing position. The possibility and effect of *Next View Planning* on *Shape from Silhouette* need to be further analyzed.

3.3 Time Schedule

The following table summarizes the steps and gives an estimate of the completion date for each step.

Description	**Estimated Completion Date**
Presentation of the thesis' specification	April 6, 2001
Implementation of automatic acquisition of multiple object views	June 15, 2001
Simple implementation of the calibration	June 30, 2001
Modification of Shape from Silhouette in order to make it independent on the calibration method used and easily extensible	July 10, 2001
Adding Shape from Structured Light to Shape from Silhouette	July 31, 2001
Design and implementation of Next View Planning	July 31, 2001
Experiments and tests	August 15, 2001
Final thesis presentation	October 1, 2001
Final diploma exam	October 15, 2001

Table 1: Time Schedule

4 Conclusion

This report proposes a new method for construction of 3D models of objects by combining *Shape from Silhouette* and *Shape from Structured Light*. The new method tries to take the best out of both underlying methods and create a more accurate and precise model of an object than it would be possible by applying only one of the methods. The implementations of the underlying methods already exist at the PRIP-Group. *Shape from Silhouette* produces a volumetric model of the object and *Shape from Structured Light* a cloud of points belonging to the object's surface. The main problem that needs to be solved is to find a 3D data structure which would be suitable for both methods in order to be able to combine them. In addition, possible ways of improving feature extraction from input images will be investigated, as well as the possibility of introduction of *Next View Planning*.

References

[AAK71] Y. I. Abdel-Aziz and H. M. Karara. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Symphosium on Close-Range Photogrammetry*, pages 1–18, Urbana, Illinois, January 1971. University of Illinois at Urbana-Champaign.

[CH88] 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.

[HS91] R. M. Haralick and L. G. Shapiro. Glossary of computer vision terms. *Pattern Recognition*, 24(1):69–93, 1991.

[Joh93] M. Johannesson. Sheet-of-light range imaging. Master's thesis, Linköping University, November 1993.

[KS99] M. Kampel and R. Sablatnig. On 3D modelling of archaeological sherds. In *Proceedings of International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, pages 95–98, 1999.

[Lis99] C. Liska. Das adaptive lichtschnittverfahren zur oberflächenkonstruktion mittels laserlicht. Technical Report PRIP-TR-55, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group, Vienna University of Technology, Austria, 1999.

[MB93] 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.

[MS96] C. Menard and R. Sablatnig. Computer based acqusition of archaeological finds: The first step towards automatic classification. In *Interfacing the Past, Computer Applications and Quantitative Methods in Archaeology*, number 28, pages 413–424, 1996.

[OTV93] C. Orton, P. Tyers, and A. Vince. *Pottery in Archaeology*, 1993.

[Pot87] 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.

[Sze93] R. Szeliski. Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58(1):23–32, July 1993.

[Tos99] S. Tosovic. Lineare Hough-Transformation und Drehtellerkalibrierung. Technical Report PRIP-TR-59, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group, Vienna University of Technology, Austria, 1999.

[Tos00] S. Tosovic. Shape from silhouette. Technical Report PRIP-TR-064, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group, Vienna University of Technology, Austria, 2000.

[Tsa86] R. Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, 1986.

[YCZB98] D. Yang, J. Chen, H. Zhou, and S. Buckley. New algorithm to calculate the center of laser reflections. In *Machine Vision Applications in Industrial Inspection VI, Proceedings of SPIE*, pages 54–58, 1998.