3-D GEOMETRY ENHANCEMENT BY CONTOUR OPTIMIZATION IN TURNTABLE SEQUENCES

Peter Eisert

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute
Einsteinufer 37, D-10587 Berlin, Germany
Email: eisert@hhi.fhg.de

ABSTRACT

A method for the enhancement of geometry accuracy in shape-from-shading frameworks is presented. For the particular case of turntable scenarios, an optimization scheme is presented that minimizes silhouette deviations which correspond to shape errors. Only three unknown parameters have to be optimized leading to a robust and relatively fast framework. In spite of the small number of parameters, experiments have shown that the silhouette error can be reduced by a factor of more than 10 even after an already quite accurate camera calibration step. The quality of an additional texture mapping can also be drastically improved making the proposed scheme applicable as a preprocessing step in many different 3-D multimedia applications.

1. INTRODUCTION

Photorealistic 3-D computer models of real objects and scenes are key components in many 3-D multimedia systems. The quality of these models often has a large impact on the acceptability of applications like virtual walk-throughs (e.g., city guides or virtual museums), caves, computer games, product presentations in e-commerce, or other virtual reality systems. Although the rendering of 3-D computer scenes can be performed in real-time even on hand-held devices, the creation and design of 3-D models with high quality is still time consuming and thus expensive. This has motivated the investigation of a large number of methods for the automatic acquisition of textured 3-D geometry models from multiple views of the object.

One class of approaches often used due to its robustness and simplicity is called *shape-from-silhouette* or *shape-from-contours* [1, 2, 3]. For shape estimation, multiple views of an object are captured and in the images, the object is segmented from the background. The contour in the image plane forms – together with the focal point of the camera – a 3-D viewing cone that contains the entire object. Intersection of all viewing cones leads to an estimate for the object geometry. In order to circumvent the limitation of reconstructing the visual hull only, many extensions have been proposed that incorporate also color information as, e.g., voxel coloring or space carving [4, 5].

Once the images are reliably segmented, shape-from-silhouette methods work very robust, are hardly affected by illumination variations, and are computationally quite efficient. With today's computers, real-time implementations now become available, enabling new applications like 3-D communication or 3-D video recording [6, 7, 8, 9]. In these scenarios, data is usually captured with multiple cameras in a cave, where the background can easily be controlled for simple segmentation.

Fig. 1. Turntable setup.

However, the accuracy of the reconstructed geometry is considerably affected by the knowledge of the true camera parameters which requires an accurate camera calibration. Deviations from the correct values can lead to smaller 3-D models, since valid object parts might be cut off during viewing cone intersection. As a result, the silhouette of the reconstructed object is always smaller or equal to the true contour. However, this unwanted effect can be used to refine the camera parameters again. In [10], Grattarola uses pairs of images and projects the viewing cone of one image contour into the other image plane, computing silhouette mismatches. Minimization of these mismatches over all camera parameters optimizes the calibration. Similarly, Niem [11] minimizes the deviation of the back-projected silhouette of the reconstructed object to the true silhouette in the camera images. Both approaches have in common, that a non-linear optimization in a high dimensional space has to be performed.

In this paper, we focus on the turntable 3-D acquisition scenario as shown in Fig. 1. Instead of placing multiple cameras around the object, a single camera is used capturing images while the object slowly rotates. The rotation angle between two shots can usually be accurately controlled, whereas the position of the camera relative to the turntable is in general unknown and requires camera calibration. In the following, we present a method for optimizing the camera's position from silhouette mismatches. In contrast to the work in [11], the circular motion adds severe constraints which increase robustness and result in a very low dimensional parameter space that has to be searched. Such constraints are also utilized in [12], where extrinsic and intrinsic camera parameters are derived from feature point correspondences instead of silhou-

0-7803-8554-3/04/$20.00 ©2004 IEEE.

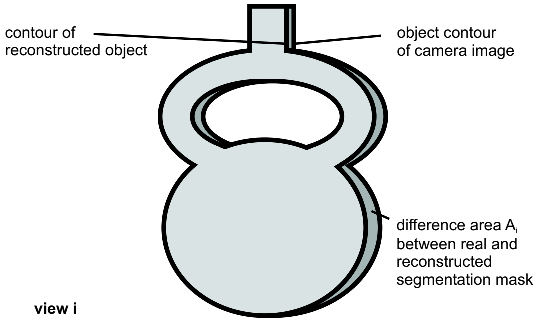

Fig. 2. Contours of the real and reconstructed object for an incorrectly calibrated camera.

ette. Experiments show that the proposed silhouette optimization converges even for large deviations making camera calibration less important or even unnecessary. Due to the improved camera parameters, the accuracy of the reconstructed 3-D models is highly increased by this preprocessing step.

2. CONTOUR OPTIMIZATION FOR TURNTABLE SETUPS

With perfectly calibrated cameras, shape-from-silhouette reconstructs the visual hull of an object by volume intersection of the viewing frustrums defined by the segmented contours. If the parameters of the real cameras are not accurately known, an erroneous object shape is reconstructed. Due to the intersection process, the reconstructed object contour always lies within the segmentation mask of the camera image as shown in Fig. 2.

The area difference A_i between correct and reconstructed object masks for view i is directly related to the amount of camera parameter deviation. Therefore, it can be used to optimize the M extrinsic and intrinsic camera parameters p_j by minimizing the area difference. Following [11], we define the cost function to be minimized as the sum of all area deviations A_i over all available N views

$$c(p_0, \ldots, p_{M-1}) = \frac{1}{N} \sum_{i=0}^{N-1} A_i(p_0, \ldots, p_{M-1}). \qquad (1)$$

In order to compute the cost function, shape-from-silhouette is performed on a voxel volume with some initial camera parameters and the reconstructed geometry is back-projected into the views creating binary object masks. The area A_i is simply defined as the number of pixels belonging to the segmented camera image but not to the reconstructed object mask. The cost function $c(p_0, \ldots, p_{M-1})$ is then minimized by searching for an optimal parameter set p_j corresponding to the correct camera parameter settings. However, in the general case of N independent views, the number M of unknown camera parameters might be very large resulting in a very tedious optimization. For the particular case of a turntable acquisition scenario, we show that the number of degrees of freedom can be drastically reduced resulting in a much faster and more robust system that can be used to obtain highly accurate 3-D models.

2.1. Parameter Constraints for Turntable Scenario

In this paper, we describe the parameter estimation and camera calibration by object contour optimization for the special case of turntable scenarios. This is particular important for off-line 3-D model acquisition for applications like e-commerce, computer games, or virtual museums. Here, accurate 3-D geometry and texture is crucial to obtain photorealistic results for arbitrary viewing directions interactively chosen by a user.

The turntable setup, however, severely restricts the number of degrees of freedom due to the use of a single camera and a perfectly circular object point trajectory. This results in robust and fast optimization compared to the general case. Since the rotation of a turntable as shown in Fig. 1 can be usually controlled accurately, the only unknowns are the extrinsic and intrinsic camera parameters. In the experiments, we assume the intrinsic parameters to be known from an initial camera calibration step. Adding the intrinsic parameters to the optimization scheme does not change the structure of the algorithm but only slightly increases the number of unknowns of the cost function. The only information that is really unknown is the position and orientation of the camera relative to the turntable. In general, this leads to 6 degrees of freedom – 3 for rotation and 3 for translation. However, for the turntable scenario we can further reduce the number of unknowns.

Assume an object point $\mathbf{x}$ rotating around the rotation axis through the point $\mathbf{x}_0$ of the turntable. For all views i, the object point $\mathbf{x}_i$ lies on a circular curve that is tilted somehow relative to the camera coordinate system

$$\mathbf{x}_i = \mathbf{R}_{axis} \cdot \mathbf{R}_i \cdot (\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0. \qquad (2)$$

In this equation, $\mathbf{R}_i$ defines the rotation around the y-axis due to turntable motion

$$\mathbf{R}_i = \begin{bmatrix} \cos R_{y,i} & 0 & \sin R_{y,i} \\ 0 & 1 & 0 \\ -\sin R_{y,i} & 0 & \cos R_{y,i} \end{bmatrix}. \qquad (3)$$

The rotation angle $R_{y,i}$ is known from the turntable control. Rotation matrix $\mathbf{R}_{axis}$ defines the orientation of the turntable rotation axis and has two unknown parameters (rotation angles R_x and R_z) that are identical for all available views

$$\begin{aligned} \mathbf{R}_{axis} &= \mathbf{R}_z \cdot \mathbf{R}_x \qquad (4) \\ &= \begin{bmatrix} \cos R_z & -\sin R_z \cos R_x & \sin R_z \sin R_x \\ \sin R_z & \cos R_z \cos R_x & -\cos R_z \sin R_x \\ 0 & 0 & \cos R_x \end{bmatrix}. \end{aligned}$$

Three other unknowns are the position of the rotation axis specified by $\mathbf{x}_0 = [t_x \quad t_y \quad t_z]^T$. However, the distance t_z of the camera to point $\mathbf{x}_0$ can be arbitrarily chosen, since this decision only scales the reconstructed object coordinates but results in the same principal shape. Similarly, the origin of the object t_y along the rotation axis does not affect the reconstructed shape.

As a result, only three unknown camera parameters have to be estimated: two rotation angles R_x and R_z for the orientation of the turntable axis and the horizontal position of the rotation axis t_x. Compared to $6 \cdot N$ unknown parameters, the optimization is much easier for this relatively low-dimensional space. For the experiments we have implemented a simple gradient descent optimization that is able to find the correct values even for very crude initial values since the cost function is relatively smooth as shown in Fig. 3.

3. EXPERIMENTAL RESULTS

In order to illustrate the performance of the proposed method for camera parameter and geometry enhancement, several exper-

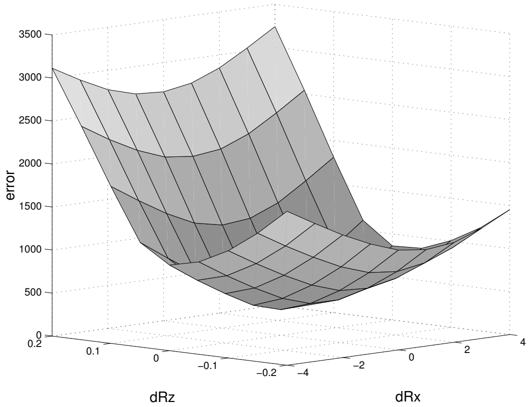

Fig. 3. Silhouette error as a function of deviations in the rotational angles R_x and R_z.

Fig. 4. Geometry enhancement for a *Moche* vase. Left: two frames of the original sequence; middle: contour error after camera calibration; right: contour error after proposed optimization.

	initial error	final error	improvement
Moche calib.	3192	271	11.8
Moche	9804	264	37.1
Tree	17461	1467	11.9

Table 1. Average error A_i before and after contour optimization.

	Δt_x	ΔR_x	ΔR_z
Moche calib.	1.5 mm	0.08^o	0.05^o
Moche	-3.4 mm	21.2^o	0.95^o
Tree	-5.7 mm	3.76^o	1.42^o

Table 2. Estimated parameter changes for the unknown parameters R_x, R_z, and t_x.

iments are performed on natural image sequences. For all sequences, every 10 degrees of turntable rotation an image is captured leading to 36 views with a resolution of 432x576 pixels. Based on background color information, the object is segmented from the background leading to 36 binary object masks. The internal camera parameters are determined once for the camera with a model-based camera calibration technique [13].

With these initial values, the optimization is started. Fig. 4 shows the results for two frames of the sequence *Moche* containing a vase from the Peruvian Moche culture. The left images depict the original frames and the middle ones the silhouette error A_i for the corresponding view after reconstruction with the parameters from the camera calibration step. The right images show the deviations between real and reconstructed contours which are significantly reduced by the proposed algorithm. The cost function (1) describing parameter inaccuracy is reduced by a factor of more than 10 as shown in the first row of Table 1.

The resulting camera parameter updates, however, are relatively small as depicted in the first row of Table 2. The initial position error of the turntable axis is about 1.5 mm and their orientation is determined with an accuracy of less than 1 degree. This indicates that even small parameter deviations from the camera calibration can lead to rather large geometry errors and optimizing these values can significantly enhance the accuracy of the reconstructed geometry.

In a second experiment, the initial values for R_x, R_z, and t_x are set to zero and no information from the previous camera calibration is exploited. The algorithm, however, still converges as indicated in Fig. 5. Although the initial silhouettes does not match, the final silhouette error is in the same range as indicated in the middle row of Table 1. The correction of the rotation angle R_x is more than 20 degrees (see Table 2).

Similar experiments are performed for a second sequence *Tree* that has a much more sophisticated contour. The algorithm still converges even for unknown initial values of the camera position and orientation. Similarly, silhouette errors befor and after optimization are depicted in Fig. 6.

The improvement of geometry accuracy resulting from the proposed method also affects the quality of the texture map estimated from the views. Incorrect shape prevents an accurate stitching of multiple texture parts from different views leading to blurred images. The left side of Fig. 7 shows the quality of the final textured 3-D model of the vase after reconstruction with parameters from the initial camera calibration. Although the refined values are quite similar, the sharpness of the texture map after the proposed geometry enhancement is drastically improved. This indicates that an additional refinement of camera parameters for shape-from-silhouette reconstruction can help to improve the quality of the results.

4. CONCLUSIONS

In this paper, we have presented a method for the refinement of camera parameters for turntable scenarios. The contour deviation between reconstructed object and segmentation mask of the camera frames is minimized in a shape-from-silhouette framework leading to an optimal set of camera parameters. For the particular case of turntable acquisition, we have shown that only 3 parameters are sufficient to calibrate the external camera parameters. In-

Fig. 5. Silhouette optimization with no initial pose information from calibration. Middle: Large initial silhouette error; right: significantly reduced deviations after optimization.

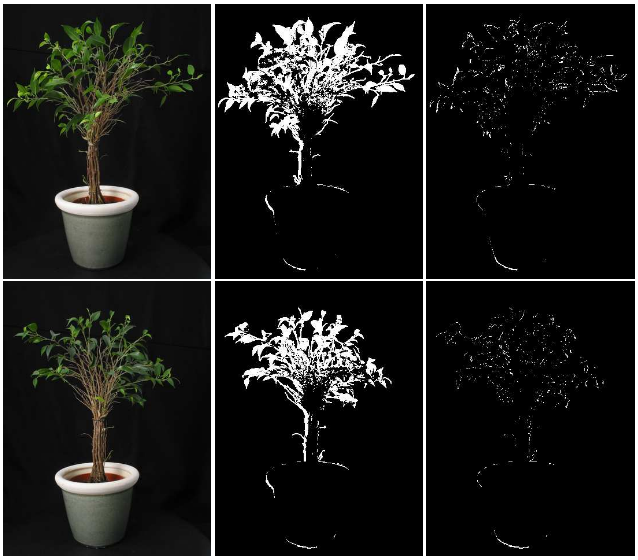

Fig. 6. Geometry deviations for the sequence *Tree*. Left: original camera frames; middle: initial error; right: silhouette error after refinement.

ternal camera parameter optimization can be added similarly without changing the algorithm. The low-dimensional parameter space ensures high robustness and reduced complexity. Experiments have shown that the quality of 3-D models can be drastically improved with the proposed optimization technique.

Fig. 7. Magnification of a reconstructed object without (left) and with (right) camera position refinement. The improved geometry results in a much sharper texture.

5. ACKNOWLEDGMENT

The work presented in this paper has been developed with the support of the European Network of Excellence VISNET (IST Contract 506946).

6. REFERENCES

[1] R. Szeliski, "Rapid octree construction from image sequences," in *Computer Vision, Graphics and Image Processing*, Jul. 1993, pp. 23–32.

[2] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150–162, Feb. 1994.

[3] W. Niem and J. Wingbermühle, "Automatic reconstruction of 3D objects using a mobile monoscopic camera," in *Proceedings of the International Conference on Recent Advances in 3D Imaging and Modelling*, Ottawa, Canada, May 1997.

[4] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," in *Proc. Computer Vision and Pattern Recognition*, Puerto Rico, 1997, pp. 1067–1073.

[5] P. Eisert, E. Steinbach, and B. Girod, "Multi-hypothesis, volumetric reconstruction of 3-D objects from multiple calibrated camera views," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Phoenix, Mar. 1999, pp. 3509–3512.

[6] X. Wu and T. Matsuyama, "Real-time active 3D shape reconstruction for 3D video," in *Proc. of 3rd International Symposium on Image and Signal Processing and Analysis*, Rome, Italy, Sep. 2003, pp. 186–191.

[7] B. Goldlücke and M. Magnor, "Real-time, free-viewpoint video rendering from volumetric geometry," in *Proc. Visual Computation and Image Processing (VCIP)*, Lugano, Switzerland, Jul. 2003.

[8] O. Grau, "Studio production system for dynamic 3D content," in *Proc. Visual Computation and Image Processing (VCIP)*, Lugano, Switzerland, Jul. 2003.

[9] M. Gross, "blue-c: A spatially immersive display and 3D video portal for telepresence," in *Proc. Computer Graphics (SIGGRAPH)*, San Diego, USA, Jul. 2003, pp. 819–827.

[10] A. A. Grattarola, "Volumetric reconstruction from object silhouettes: A regularization procedure," *Signal Processing*, vol. 27, pp. 27–35, 1992.

[11] W. Niem, *Automatische Rekonstruktion starrer dreidimensionaler Objekte aus Kamerabildern*, Ph.D. thesis, University of Hanover, Hanover, Germany, 1998.

[12] A. W. Fitzgibbon, G. Cross, and A. Zisserman, "Automatic 3D model construction for turn-table sequences," in *Proc. ECCV 98 Workshop on 3D Structure from Multiple Images in Large-Scale Environments*, Freiburg, June 1998, pp. 154–170.

[13] P. Eisert, "Model-based camera calibration using analysis by synthesis techniques," in *Proc. Vision, Modeling, and Visualization VMV'02*, Erlangen, Germany, Nov. 2002, pp. 307–314.