Error Analysis for Silhouette–Based 3D Shape Estimation from Multiple Views

Wolfgang Niem

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany
niem@tnt.uni–hannover.de http://www.tnt.uni–hannover.de/~niem

ABSTRACT

This paper presents an error analysis for 3D shape estimation techniques using silhouettes from multiple views (”shape–from–silhouette”). The results of this analysis are useful for the integration of other shape reconstruction techniques into the shape–from–silhouette approach and the appropriate choice of its parameters. The analysis allows an assessment of the local 3D shape error and takes errors into account which result from the finite resolution of the silhouettes and the volume, the finite number of views, as well as from the influence of erroneous camera parameters.

1. INTRODUCTION

3D models of real 3D objects are the components of virtual worlds in a variety of 3D multimedia applications like 3D teleshopping, virtual studio production, 3D teleconferencing, 3D information systems, 3D archiving, etc.. The manual reconstruction of those 3D objects is time consuming and therefore cost expensive. For that reason, techniques are under investigation that allow the automatic reconstruction of real 3D objects.

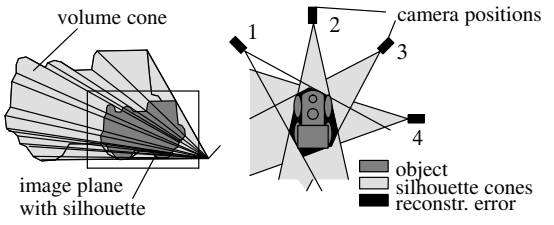


Fig. 1 shape–from–silhouette: (a) construction of a volume cone (b) intersection of volume cones (2D top view)

A robust approach for the automatic reconstruction of the 3D shape using multiple camera views is the ”shape–from–silhouette” technique [4][6][9], which can be divided into the following steps: First, the silhouettes are extracted from the input images, for example by color keying. Then, for each silhouette a volume cone is reconstructed in a second step (Fig. 1a). Therefore, the camera parameters are estimated, e.g. by using an underlying planar calibration pattern [5]. Finally, the intersection of the volume cones results in the bounding volume (Fig. 1b).

The main characteristics of this technique are:

+ robust shape estimation due to simple and accurate silhouette detection
+ unique 3D–data representation (voxel representation) even for objects with a complex surface topography
+ high precision for convex object regions
– inability for the estimation of concave surfaces

Two reasons make it desirable to have an estimate of the local 3D shape error for the 3D models :

(i) For the estimation of concave surfaces, the combination of silhouette–based shape estimation with other shape estimation techniques e.g. using stereo images, laser triangulation or structured light [2][8] is suitable. For a proper combination of the different shape estimators by using a minimum variance estimator, the individual error variances are required.

(ii) The choice of the number of views and the voxel resolution of the volume are free parameters of the 3D shape reconstruction process. Hence, it is useful to support the user to adjust those parameters properly with respect to the silhouette resolution and the accuracy of the camera parameters.

Therefore, this paper presents an analytical error analysis for silhouette–based 3D shape estimation, which results in an estimate of the local shape error. Specifically, shape errors are taken into account which result from the spatial quantization of the silhouette images, the spatial quantization of the volume representation, the limited number of views and erroneous camera parameters. In section 2 the used camera model is defined, section 3 introduces the shape error measure. In section 4 the different errors are analyzed and section 5 presents some results.

2. CAMERA MODEL

For the presented error analysis, the pinhole camera model as defined e.g. by Yakimovski [8] is used. For a more detailed modelling of the imaging errors introduced by real CCD cameras, the study of Lenz and Fritsch [3] is recommended.

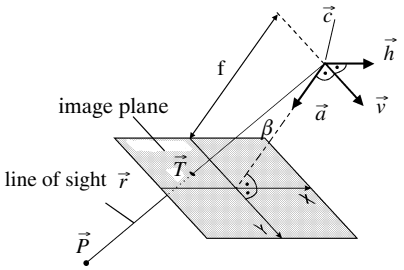


Fig. 2 Projection of a real–world point $\vec{P}$ into an image point $\vec{p} = (I,J)$

Denoting

$\vec{c}$: focal point
$\vec{a}$: unit vector in the direction of the optical axis
$\vec{h}$: unit vector parallel to the horizontal axis of the image plane
$\vec{v}$: unit vector parallel to the vertical axis of the image plane
f : focal length of the camera

the projection (Fig. 2) of a point $\vec{P}$ along the line of sight $\vec{r}$ into the image point $\vec{p} = (I,J)$ can be performed using

$$I = \frac{(\vec{P} - \vec{c}) \cdot \vec{h}}{(\vec{P} - \vec{c}) \cdot \vec{a}} \cdot f \quad ; \quad J = \frac{(\vec{P} - \vec{c}) \cdot \vec{v}}{(\vec{P} - \vec{c}) \cdot \vec{a}} \cdot f \qquad (1)$$

Furthermore, the distance of a point $\vec{P}$ from the focal point perpendicular to the sensor image plane is denoted with z and the angle between the optical axis $\vec{a}$ and the line of sight $\vec{r}$ is denoted with β.

3. SHAPE ERROR MEASURE

As the measurements from all viewpoints are integrated in the 3D voxel space (Fig. 3a) (which is equivalent to the pixel space of an image), a shape error measure is defined, which allows the quantitative acquisition of the different error sources locally.

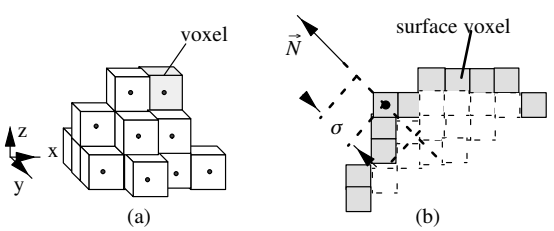


Fig. 3 (a) voxel representation of a volume (b) 2–D cross section of a voxel representation with σ indicating the standard deviation of the position error in the direction of the surface normal $\vec{N}$ of a surface voxel.

For this criterion, the position error is evaluated in the *direction of the surface normal* $\vec{N}$ (further referred to as DSN error) for each surface voxel of the volume model (Fig. 3b). All the investigated error sources are evaluated with respect to the best convex approximation of the volume ("visual hull").

4. ERROR ANALYSIS

The different shape reconstruction errors relevant to the shape–from–silhouette approach can be classified roughly into three categories:

(i) errors resulting from position errors of feature points (i.e. silhouette contour points) in the image plane. This accuracy depends on image processing operations used for feature point localisation (i.e. image segmentation), as well as the accuracy of the camera parameters and the quality of the camera model used. In this paper, the spatial quantization of image segmentation and erroneous camera parameters are discussed.

(ii) errors resulting from 3D operations in the 3D space. In this paper, the quantization error of the 3D position in the voxel space is investigated.

(iii) errors resulting from the number and positions of evaluated views.

4.1. Quantization of silhouette images

The segmentation of the silhouettes is obtained in the current approach by using colour keying and is represented in a binary image with picture elements (pel) of the size $s_x \cdot s_y$. The position of pels on the edge of the silhouette (silhouette contour points), which is most important for the shape reconstruction is distorted due to spatial image quantization.

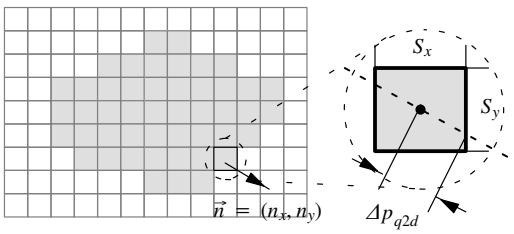


Fig. 4 Binary silhouette image with picture elements of the size $s_x \cdot s_y$ and the maximal position error Δp_{q2d} of a silhouette contour point in the direction of the silhouette contour normal $\vec{n} = (n_x, n_y)$

The maximal position error due to quantization Δp_{q2d} *in the direction of the silhouette contour normal* $\vec{n} = (n_x, n_y)$ (Fig.4) is crucial for the resulting maximal 3D shape DSN error ΔP_{Q2D} and can be calculated by

$$\Delta p_{q2d} = \frac{1}{2} \sqrt{{s_x}^2 + \left(\frac{n_y}{n_x} s_y\right)^2} \quad , |n_x| \geq |n_y| \qquad (1)$$

$$\Delta p_{q2d} = \frac{1}{2} \sqrt{{s_y}^2 + \left(\frac{n_x}{n_y} s_x\right)^2} \quad , |n_y| \geq |n_x| \qquad (2)$$

The maximal 3D shape DSN error ΔP_{Q2D} can be calculated by using the camera arrangement (Fig. 2)

$$\Delta P_{Q2D} = \frac{\Delta p_{q2d}\ z\ \cos\beta}{f} \tag{3}$$

Assuming an equally distributed position error in $[-\Delta p_{q2d},\ \Delta p_{q2d}]$, the variance of the 3D shape DSN error resulting from silhouette quantization can be calculated by

$$\sigma^2_{Q2D} = \frac{(2\Delta P_{Q2D})^2}{12} = \frac{(\Delta p_{q2d}\ z\ \cos\beta)^2}{3\ f^2} \tag{4}$$

4.2. Accuracy of camera parameters

The accuracy of camera parameters also results in a 2D position error in the image plane. A calibration technique e.g. as introduced by Tsai [7] evaluates the camera parameter set $\{\vec{c}, \vec{a}, \vec{h}, \vec{v}, f\}$ by minimizing the sum of the squared differences E_i between N extracted 2D calibration points $p_{localised\,i}$ and the projections $p_{cahvf\,i}$ (using eq. (1)) of the N corresponding 3D points of a known calibration pattern:

$$\sum_{i=1}^{N} |\ E_i\ |^2 \xrightarrow[\vec{c}, \vec{a}, \vec{h}, \vec{v}, f]{} \min \tag{5}$$

The standard deviation of the normally distributed residual errors in the image plane is given by

$$\sigma_{cam_2d} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} |\ p_{localised\,i} - p_{cahvf\,i}|^2} \tag{6}$$

and analogue to eq. (3) and (4) the resulting variance of the DSN error can be calculated by

$$\sigma^2_{CAM_3D} = \frac{(\sigma_{cam_2d}\ z\ \cos\beta)^2}{f^2} \tag{7}$$

This result has to be used carefully regarding to the kind of calibration pattern used: If a planar pattern is used, eq. (7) is only exactly valid within the 3D plane defined by the calibration pattern. This results from the fact, that the calibration is not unique in this case. Thus, a good mapping of 3D points lying in the calibration plane is obtained, but the mapping may become worse for 3D points, which are distant from this plane.

A sensibility analysis of the single camera parameters using eq. (1) can give information about critical parameter combination but goes beyond the scope of this paper.

4.3. Quantization of the volume

The 3D shape DSN errors resulting from the spatial quantization of the volume depend on the user defined resolution of the voxel representation.

Depending on the side length a of a surface voxel and its normal direction $\vec{N} = (N_x, N_y, N_z)$ (Fig. 5) the maximal DSN error ΔP_{Q3D} is given by

$$\Delta P_{Q3D} = \frac{a}{2}\sqrt{\left(\frac{N_y}{N_x}\right)^2 + \left(\frac{N_z}{N_x}\right)^2 + 1}\ ,\ |N_x| \geq |N_y|, |N_x| \geq |N_z|$$

$$\Delta P_{Q3D} = \frac{a}{2}\sqrt{\left(\frac{N_x}{N_y}\right)^2 + \left(\frac{N_z}{N_y}\right)^2 + 1}\ ,\ |N_y| \geq |N_x|, |N_y| \geq |N_z|$$

$$\Delta P_{Q3D} = \frac{a}{2}\sqrt{\left(\frac{N_x}{N_z}\right)^2 + \left(\frac{N_y}{N_z}\right)^2 + 1}\ ,\ |N_z| \geq |N_x|, |N_z| \geq |N_y|$$

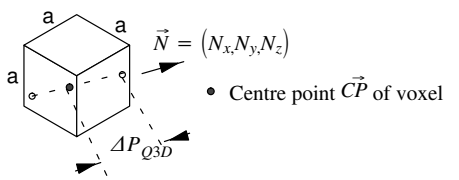


Fig. 5 Voxel with side length a, the voxel surface normal $\vec{N}$ and the maximal DSN error ΔP_{Q3D}

Assuming an equally distributed 3D position error in $[-\Delta P_{Q3D},\ \Delta P_{Q3D}]$, the variance of the 3D shape DSN error is given by

$$\sigma^2_{Q3D} = \frac{\Delta P_{Q3D}^{\ 2}}{3} \tag{8}$$

4.4. Number of views

The number and choice of views is a crucial point in the shape–from–silhouette approach. In contrary to the other errors, the DSN error resulting from the finite number of views is a *systematic* one. As shown in Fig. 1b, the estimated volume is always larger than the real volume and therefore, the DSN error is always negative.

Assuming objects with a smooth surface and a locally constant radius of curvature *r*, the resulting 3D DSN error depends on the angle $\Delta\phi$ between two consecutive viewpoints and the radius *r* of curvature.

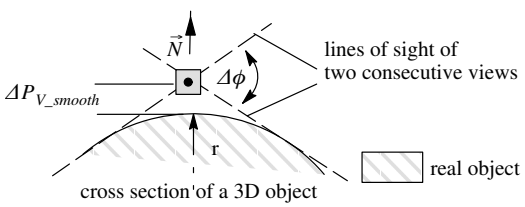


Fig. 6 Illustration of the maximal DSN error ΔP_{V_smooth} assuming a local radius *r* of curvature

The maximum 3D DSN error ΔP_{V_smooth} as defined in Fig. 5 can be calculated by:

$$\Delta P_{V_smooth} = r\left(\frac{1}{\cos(\frac{\Delta\phi}{2})} - 1\right) \approx r\frac{(\Delta\phi)^2}{8}, \Delta\phi < 30^o \tag{9}$$

It turns out, that the error grows linearly with the radius *r* of curvature and quadratically with the angle $\Delta\phi$ between two views. For surface regions with a low curvature the local reconstruction error is larger compared to those for surface regions with high curvature. The radius *r* can be estimated from three consecutive views [9].

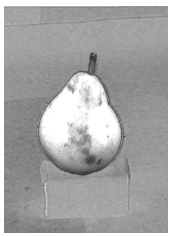
Reducing the model assumption to arbitrary convex objects, from geometrical constraints it can be shown, that the maximal DSN error $\Delta P_{V_arbitrary}$ can be evaluated from $\Delta\phi$ and the estimated "radius" r to

$$\Delta P_{V_arbitrary} \approx r\frac{(\Delta\phi)^2}{2}, \quad \Delta\phi < 30^o \tag{10}$$

Even if this error is a systematic one (as from geometrical constraints the *direction* of the error is known) a meaningful compensation is not possible and therefore the DSN error $\Delta P_{V_arbitrary}$ is added as a signed measurement uncertainty to the measured surface point.

5. RESULTS

The potential of the presented analysis is exemplary shown by its application to the real input sequence "pear" (Fig.7) . For this example, the voxel side length a and the number of viewpoints N should be chosen such, that they cause maximal shape errors in the same order of magnitude than those caused by spatial quantization of the silhouette images. The quantization error is fixed by the parameters given in Fig. 7.



CAMERA

pixel size sx	: 0.01 mm
pixel size sy	: 0.01 mm
focal length f	: 28 mm
distance z	: 1175 mm
angle β	: $\approx 2^o$

OBJECT

estimated maximal radius of curvature r : $\approx$ 32 *mm*

Fig. 7 Input image of analyzed sequence "pear" and related parameters

Combining eq. (3) and (8), and taking the worst case $\Delta P_{Q3D} = \frac{a}{2}\sqrt{3}$ and $\Delta p_{q2d} = \frac{1}{2}s_x$, the proposed voxel side length a should be chosen such that:

$$a = \frac{s_x\, z\, \cos\beta}{\sqrt{3}\, f} \approx 0.24\ mm \tag{11}$$

For equally distributed viewpoints (around a fixed axis of rotation) and combining eq. (3) and (9), the number of views $N = \frac{2\pi}{\Delta\phi}$ should be chosen as:

$$N \geq \pi \sqrt{\frac{r\, f}{s_x\, z\, \cos\beta}} \approx 27 \tag{12}$$

Those analytically obtained values for N and a coincide well with empirically found values.

6. SUMMARY

An error analysis for 3D shape estimation techniques using silhouettes from multiple views has been presented. As local shape measure, the position error of a surface voxel of the volume model in the direction of the surface normal has been evaluated for different error sources:

(i) quantization of the silhouette images, which is determined by the camera parameters,

(ii) accuracy of the camera parameters, which is determined by the calibration technique applied,

(iii) quantization of the volume, which can be influenced by the user through the choice of the voxel space resolution, and

(iv) the number of viewpoints, which can be influenced by the user and depends on the measured local curvature.

The analysis has been successfully applied to a real image sequence in order to obtain an appropriate choice for the voxel space resolution and the number of views.

In future work, the shape–from–silhouette approach with analyzed local shape errors will be combined with other shape estimation techniques e.g. by using a minimum variance estimator.

7. REFERENCES

[1] Cyberware, Monterey, CA 93940, http://www.cyberware.com/

[2] L. Falkenhagen, "Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuos Surfaces", published in "Image Processing for Broadcast and Video Production", Springer Verlag, ISBN 3–540–19947–0.

[3] R. Lenz, D. Fritsch, "Accuracy of videometry with CCD sensors", *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 45, No. 2, pp 90 – 110, 1990.

[4] W. Niem, "Robust and Fast Modelling of 3D Natural Objects from Multiple Views", *SPIE Proceedings "Image and Video Processing II"*, Vol. 2182, 1994, pp. 388–397.

[5] W. Niem, J.Wingbermühle: "Automatic Reconstruction of 3D Objects Using a Mobile Monoscopic Camera", *IEEE Proceedings of the International Conference on Recent Advances in 3D Imaging and Modelling*, Ottawa, Canada 12–15 May 1997.

[6] R. Szeliski, "Rapid Octree Construction from Image Sequences", *CVGIP: Image Understanding*, Vol. 58, No 1, July, pp. 23–32, 1993.

[7] R. Y. Tsai, "A Versatile Camera Calibration Technique for High–Accuracy 3D Machine Vision Metrology Using Off–the–Shelf TV Cameras and Lenses", *Journal of Robotics and Automation*, Vol. RA–3. No.4, August 1987, pp. 323–344.

[8] Y.Yakimovsky, R. Cunningham, "A System for Extracting Three–Dimensional Measurements from a Stereo Pair of TV Cameras", *Computer Vision, Graphics, and Image Processing*, Vol. 7, pp. 195 –210, 1978.

[9] J.Y. Zheng, "Acquiring 3–D Models from Sequences of Contours", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 16, No. 2, February 1994.