Automatic Reconstruction of 3D Objects Using a Mobile Monoscopic Camera

Wolfgang Niem, Jochen Wingbermühle

Universität Hannover
Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung
Appelstr. 9A, 30167 Hannover

Abstract

A method for the automatic reconstruction of 3D objects from multiple camera views for 3D multimedia applications is presented. Conventional 3D reconstruction techniques use equipment that restrict the flexibility of the user. In order to increase this flexibility, the presented method is characterized by a simple measurement environment, that consists of a new calibration pattern placed below the object allowing object and pattern acquisition simultaneously. This ensures, that each view can be calibrated individually. From these obtained calibrated camera views, a textured 3D wireframe model is estimated using a shape–from–silhouette approach and texture mapping of the original camera views. Experiments with this system have confirmed a significant gain of flexibility for the user and a drastic reduction of costs for technical equipment while ensuring comparable model quality as conventional reconstruction techniques at the same time.

1 Introduction

Natural looking 3D models of real 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 creation of those 3D models is time consuming and therefore cost expensive. For that reason, techniques are under investigation that allow the automatic reconstruction of 3D objects. Those techniques can be subdivided into two kinds of methods. Active methods use structured light or a laser scanner [12]. Passive methods use image views taken by a camera. The presented approach belongs to the passive methods which require less equipment and can be applied more generally.

A robust technique for the reconstruction of 3D objects using multiple calibrated camera views was proposed by Busch [1]. In a first step "shape from silhouettes" [2][3][11] is used for the reconstruction of the 3D shape. The resulting volume model is approximated by a wireframe model in a second step and in a third step the real object texture [4] is projected onto this wireframe model to obtain a natural look.

For the acquisition of the camera views systems consisting of a turntable with the real object precisely rotating in front of a stationary camera are used presently. Alternatively, the camera is rotated around the object. The calibration of the camera is performed in an additional processing step by evaluation of images of a calibration pattern placed on the turntable.

The application of such systems is restricted by

- the costs for the technical equipment (especially the turntable)
- the fixed position of the camera with respect to the rotational plane of the turntable
- the inability to reconstruct installed objects
- the object size
- the fixed focus of the used camera

For this reason, in this paper a novel method is presented, which is characterized by a simple measurement environment that allows a free movement of the camera around the object and an independent choice of the focus for each camera view. This is achieved by simultaneous acquisition of object and a newly developed calibration pattern.

In Section 2 the developed measurement environment is defined in detail. Section 3 describes the camera calibration, specifically the segmentation of the calibration pattern and the extraction of the calibration points. Section 4 presents the processing pipeline from the calibrated camera views to the textured wireframe model. Finally, in section 5 results for real camera views are given and discussed.

2 Measurement environment

The measurement environment necessary for the automatic object reconstruction should allow image acquisition with a mobile camera (e.g. a CCD camera or a reflex camera in combination with a slide scanner) and should be as simple as possible in view of the equipment costs. Specifically, a free camera movement around the object and arbitrary focal length for each view should be possible. For this purpose, each camera view must be calibrated with respect to a world coordinate system. This is achieved in the presented approach by simultaneous acquisition of object and calibration pattern.

Therefore, the calibration pattern must meet the following criteria:

- Object, calibration pattern and background must be separable
- Simple extraction of calibration points
- Occlusions of the object by the pattern must be prevented
- Object and calibration pattern must be in depth of focus
- Calibration with partly occluded pattern must be possible

The proposed measurement environment is shown in Fig.1. A rotationally symmetrical calibration pattern is arranged around a 3D object so that concurrent acquisition of object images and calibration images is enabled.

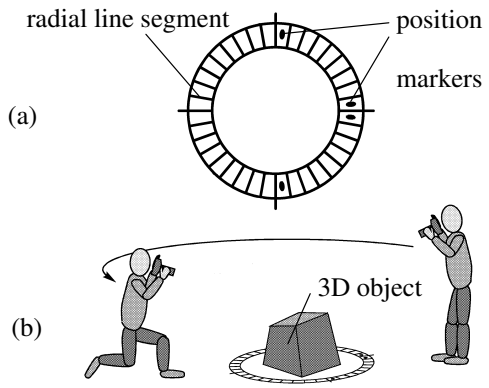


Fig. 1 Calibration pattern (a) and its usage within the measurement environment (b)

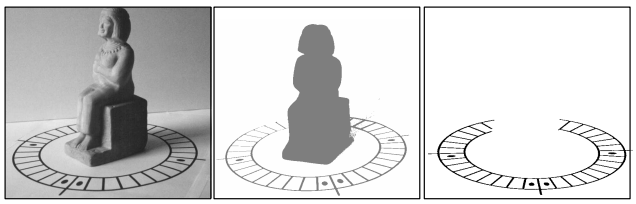
The pattern consists of two concentric circles connected by radial line segments with known dimensions. Together with position markers in each quadrant, a unique assignment of the camera views to a fixed world coordinate system is possible, even if parts of the pattern are occluded. Furthermore, this arrangement ensures that both, object and calibration pattern, are within the depth of focus and that occlusion of the object by the pattern is prevented. The background of the scene and the calibration pattern is of one colour in order to facilitate the segmentation.

3 Camera calibration

The applied camera calibration process can be subdivided into three step. In a first step, the calibration pattern is separated from object and background in the input image. The calibration points are extracted from the image and identified in a second step and in the last step the camera parameters (i.e. external parameters like position and orientation of the camera as well as internal parameters like focal length and radial distortion) are estimated using an adequate camera model.

3.1 Segmentation of calibration pattern

The separation of the calibration pattern in the input images from background and object is performed in the following steps: In a first step colour keying is used to separate the calibration pattern together with the object from the background. This results in the binary foreground mask as shown in Fig 2b. For the separation of object and calibration pattern in a second step it is assumed that the object silhouette is a single large connected region, while the pattern consists of several thin lines. A repeated erosion–dilatation step eliminates the thin lines of the pattern and generates a rough mask for the object. This mask is used to blend out the object from the foreground mask. Finally, the largest connected foreground region is extracted as the pattern mask which is the visible part of the calibration pattern (Fig 2c).



(a) input image (b) foreground mask (c) pattern mask

Fig. 2 Segmentation of the calibration mask; colour keying of the original image (a) leads to the foreground mask (b) which is used to extract the pattern mask (c)

Alternatively, the separation of object and calibration pattern could be performed by using an additional colour segmentation step as proposed in [8] for camera calibration in virtual studio environments. Therefore, background and calibration pattern had to be fabricated in the same colour but different shades of blue. This is applica-

ble to a wider class of objects and will be implemented in future systems.

3.2 Extraction of calibration points

From the pattern mask now the calibration points have to be extracted and identified. For a precise localisation of the calibration points, it is favourable to calculate the intersection of the projected radial line segments and the two concentric circles of the calibration pattern (ellipses) analytically. Therefore, the parameters of the line segments and the two ellipses have to be estimated.

For the identification of the calibration points, at least one position marker and its two neighbouring line segments have to be detected. This ensures a clear assignment of all of the projected line segments in the image plane to the corresponding line segments of the pattern.

The following processing steps are implemented:

Extracting the sample points of the ellipses and projected line segments

First, the segment regions within the pattern mask are segmented and labelled and an ellipse through the centre of gravity of the segment regions is roughly estimated (Fig. 3). Starting from the ellipse center, the pattern mask is scanned in radial direction and the sample points of the inner ellipse and the outer ellipse are detected. (Fig.4a).

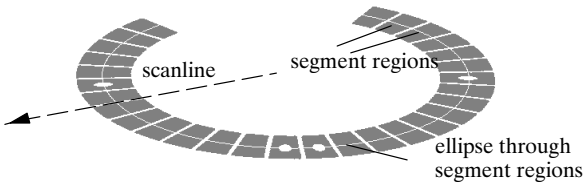


Fig. 3 Ellipse through the centers of segment regions and radial scanline used for the extraction of the sample points of the ellipse and the projected line segments

Similarly, the position markers are detected on scanlines between the inner and the outer ellipse. Both, the sample points of the ellipses as well as of the position markers are substracted from the original calibration mask and results in the sample points of the projected line segments (Fig.4b). The evaluation of the position markers allows a sorting of the projected line segments.

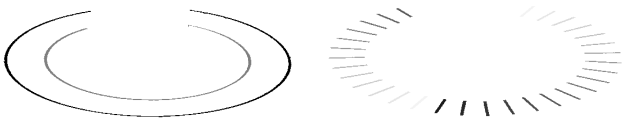


Fig. 4 (a) sample points of the ellipses
(b) sample points of the projected line segments

Estimation of the two ellipses

For the estimation of the ellipses from the extracted sample points a two–step–approach is applied. At first a linear estimation technique is used to estimate the parameter k_i of a polynomal representation of the ellipse in the form

$$k_1X^2 + 2k_2XY + k_3Y^2 + 2k_4X + 2k_5Y + k_6 = 0 \quad (1)$$

No start value is necessary for this estimation step. In order to ensure the estimation of a unique polynomal representation, at least an arc length of $120°$ must be visible in the pattern mask.

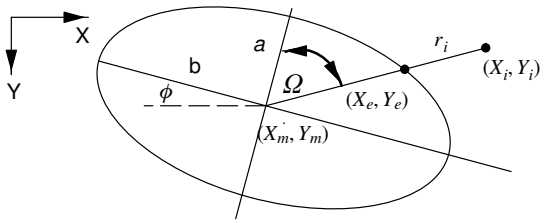


Fig. 5 Parameters of an ellipse in arbitrary situation:
(X_e, Y_e) : ellipse point
(X_m, Y_m) : centre point
a,b : length of half axes
Φ : angle of half axes with respect to coordinate axes

This estimate is improved in terms of robustness against outliers and accuracy by using it in a second step as start value for a nonlinear estimation. For this purpose the ellipse is represented in the parameter form with Ω being used for the ellipse angle:

$$\begin{aligned} X_e &= a\cos(\phi)\cos(\Omega) - b\sin(\phi)\sin(\Omega) + X_m \\ Y_e &= a\sin(\phi)\cos(\Omega) + b\cos(\phi)\sin(\Omega) + Y_m \end{aligned} \quad (2)$$

Now the sum of the squared radial distances r_i between the sample points (X_i, Y_i) and the ellipse points (X_e, Y_e) is minimized with respect to the half axes a and b, the angle Φ and the ellipse center (X_m, Y_m) (Fig. 5).

$$\sum_{i=1}^{K} |\, r_i \,|^2 \xrightarrow[a,b,\phi,X_m,Y_m]{} \min! \quad (3)$$

Therefore, r_i is determined by

$$r_i = \sqrt{\tilde{X}_i^2 + \tilde{Y}_i^2}\left(1 - \frac{a\,b}{\sqrt{\tilde{X}_i^2\,b^2 + \tilde{Y}_i^2\,a^2}}\right) \quad (4)$$

with

$$\tilde{X}_i = \sin(\Phi)\,(Y_i - Y_m) + \cos(\Phi)\,(X_i - X_m) \quad (5)$$

$$\tilde{Y}_i = \cos(\Phi)\,(Y_i - Y_m) - \sin(\Phi)\,(X_i - X_m) \quad (6)$$

For the minimization of this nonlinear sum of squared errors the Davidson–Fletcher–Powell algorithm [5] is used. The resulting ellipses are shown in Fig. 6.

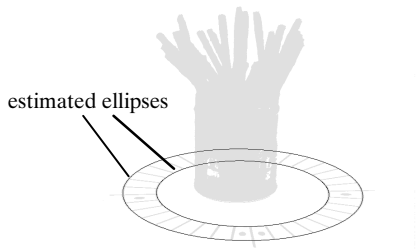


Fig. 6 Estimated ellipses blended into the foreground mask

Estimation of the line segments

For the estimation of the parameters of the radial lines the points of opposite line segments as well as the projected center point of the pattern are used. Representing the lines in the form

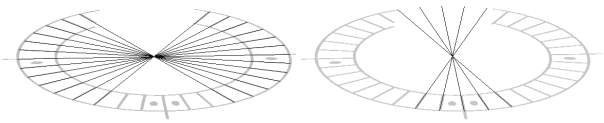
$$Y_l = m\ X_l + b \tag{7}$$

a robust estimation is performed by a minimization of the nonlinear sum of the squared distances d_i of the samples (X_i, Y_i) to the line point (X_l, Y_l) with respect to the gradient m and the section of the y–axis b.

$$\sum_{i=1}^{K} |\ d_i\ |^2 \xrightarrow[m,b]{} \text{min!} \tag{8}$$

Therefore, d_i is determined by

$$d_i = \frac{m\ X_i - Y_i + b}{\sqrt{m+1}} \tag{9}$$



(a) (b)

Fig. 7 Estimated lines blended into the pattern mask
(a) line segments with opposite counterparts
(b) line segments without opposite counterparts

For the minimization again the Davidson–Fletcher–Powell algorithm [5] is used. The resulting estimated lines are blended into the pattern mask and are shown in Fig. [7]

Calculation of intersections between line segments and ellipses

The calibration points are analytically determined by calculating the intersection between the estimated line segments and ellipses. For this purpose the parametric representation of the ellipses (Eq. (2)) is put in the equation of the line segments (Eq. (7)). The calculated calibration points are shown in Fig. 8 and can now be used together with the 2D–3D correspondence information for camera parameter estimation.

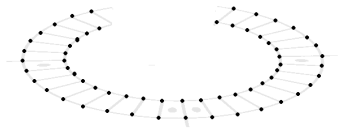


Fig. 8 Extracted calibration points resulting from the intersection of estimated ellipses and line segments

3.3 The camera model and parameter estimation

For camera calibration, the real camera must be represented by a mathematical model, which describes the imaging process with an adequate precision. A camera model, which is suited for 3D reconstruction applications, is the "cahv"–model introduced by Yakimovski and Cunningham [10].

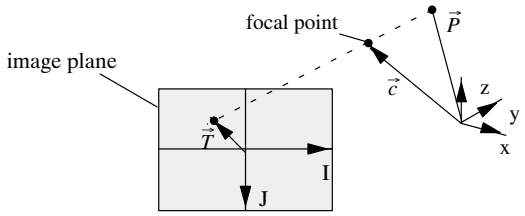


Fig. 9 Projection of a real–world point $\vec{P}$ into an image point $\vec{T} = (I,J)$

Denoting

$\vec{c} = (c_x, c_y, c_z)^T$	: vector to the focal point
$\vec{a} = (a_x, a_y, a_z)^T$	: unit vector in the direction of the optical axis
$\vec{h} = (h_x, h_y, h_z)^T$	: unit vector parallel to the horizontal axis of the image plane
$\vec{v} = (v_x, v_y, v_z)^T$	: unit vector parallel to the vertical axis of the image plane
f :	: focal length of the camera

the projection (Fig. 9) of a point $\vec{P}$ into the image point $\vec{T} = (I,J)$ can be performed using

$$I = \frac{(\vec{P}-\vec{c})\cdot\vec{h}}{(\vec{P}-\vec{c})\cdot\vec{a}}\cdot f \quad ; \quad J = \frac{(\vec{P}-\vec{c})\cdot\vec{v}}{(\vec{P}-\vec{c})\cdot\vec{a}}\cdot f \tag{10}$$

This model allows a linear distortion and a shift of the optical center with respect to the image center.

3rd order radial distortions are compensated with this model using

$$I = \frac{I_d}{1 + k_r r_d^2}; \; J = \frac{J_d}{1 + k_r r_d^2};$$

$$r_d^2 = I_d^2 + J_d^2 \tag{11}$$

The camera parameter set $\{\vec{c}, \vec{a}, \vec{h}, \vec{v}, f, k_r\}$ is estimated, by minimizing the sum of the squared differences E_i between the N extracted 2D calibration points and the projections of the N corresponding 3D points of pwdthe pattern:

$$\sum_{i=1}^{N} | E_i |^2 \xrightarrow[\vec{c}, \vec{a}, \vec{h}, \vec{v}, f, k_r]{} \text{min!} \tag{12}$$

This parameter estimation can be performed by using a calibration technique like the one proposed by Tsai [7].

For an assesment of the obtained estimation accuracy with respect to the resolution of the used camera sensor, Weng [9] introduced the normalized calibration error

$$n_{ce} = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N} | E_i |^2}{\sigma_q^2}} \tag{13}$$

with σ_q^2 being the variance of the equally distributed quantization noise of a camera sensor with a pixel size of $S_X \; x \; S_Y$:

$$\sigma_q^2 = \frac{S_X^2 + S_Y^2}{12} \tag{14}$$

A normalized calibration error $n_{ce} \leq 1$ indicates a good calibration.

4 3D reconstruction

The automatic 3D reconstruction of the real objects can be subdivided into the shape reconstruction part resulting in a wireframe model and the texture mapping, which assigns a natural look to the model.

4.1 Shape reconstruction

One approach for the reconstruction of 3D objects using silhouettes from multiple camera views is called "shape from silhouettes "or "method of occluding contours" [2][6]. The principle of this approach can be divided into two steps.

In a first step, a bounding pyramid is constructed using the focal point of the camera and the silhouette (Fig. 10c). The convex hull of this pyramid is formed by the lines of sight from the camera focal point through different contour points of the object silhouette.

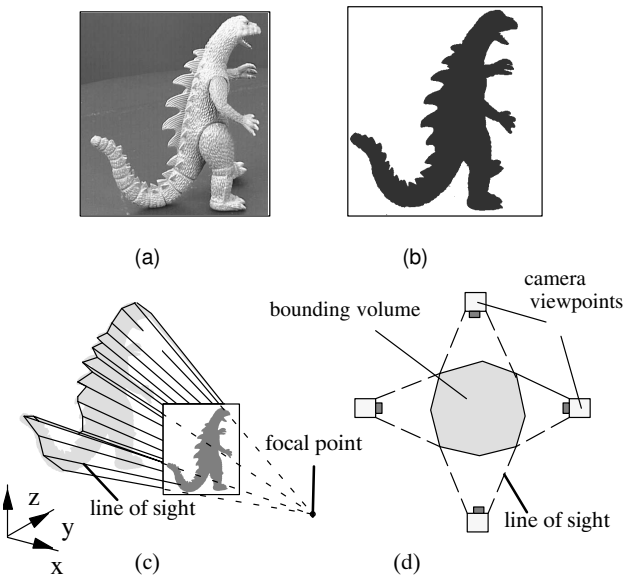


Fig. 10 Shape from silhouettes
(a) original image, (b) silhouette
(c) bounding pyramid
(d) topview: intersection of bounding pyramids

In a second step, the pyramids are intersected and form an approximation of the bounding volume (Fig. 10d).

The synthesis algorithms used in computer animation work with surface models. Thus, the volume representation must be transformed into a surface model. For that purpose the surface of the volume model is approximated by a triangle mesh. A mesh growing algorithm is used which allows a local adaptation of the volume model surface by adapting the size of each triangle patch according to a tolerable approximation error. This results in smaller triangles in surface regions with high curvature and larger triangles in surface regions with low curvature.

Starting with a single triangle placed on the volume model surface further triangles are constructed at each open edge until the whole volume model is covered with a mesh (Fig. 11).

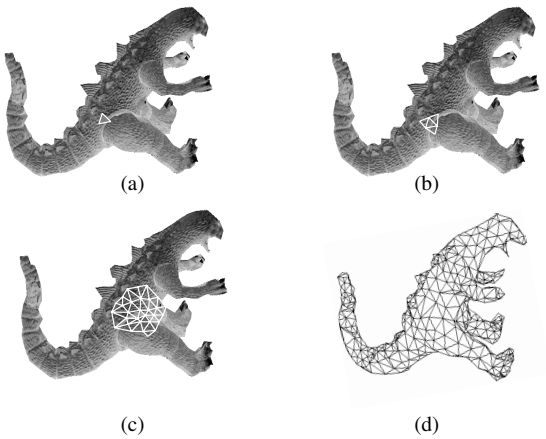


Fig. 11 Approximating the volume model by a triangle mesh; (a) starting triangle, (b) neighbouring triangles, (c) propagation of the mesh, (c) closing of the mesh

4.2 Texture mapping

For natural looking 3D models the texture is of great importance. To meet this goal a texturing algorithm is used which estimates the texture for each surface triangle from the input image sequence [4].

The principle of binding texture information to a single surface triangle is explained in Fig. 12. By using the cahv–camera parameters the vertices of a surface triangle are projected into the camera plane with the original image. The clipped rectangular image part containing the projected triangle is defined as texture map.

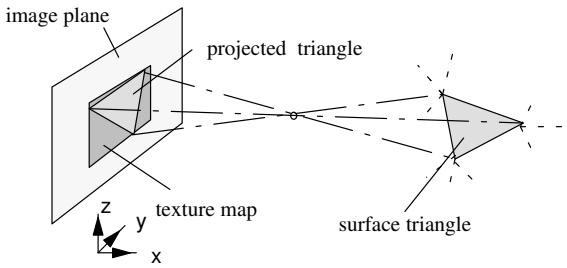


Fig. 12 Binding a texture map to a surface triangle

The quality of the final texture depends on illumination and the resolution of the CCD camera. Furthermore the correct binding of the texture to the 3D surface model is influenced by the accuracy of camera position relative to the object and by shape errors. All these influencing factors have to be taken into account in the texturing process in order to obtain a model with low distortion of texture.

The used texturing method is divided into three steps:

Grouping triangles to surface regions textured from a common image

As the texture at boundaries between surface regions textured from different images is probably distorted, those regions should be as homogeneous as possible in order to reduce the total length of boundaries. Furthermore, the assignment of a triangle to a surface region depends on its texture resolution, which is defined as the ratio of pixel elements per surface unit and should be tolerable within a surface region.

Local texture filtering of the boundaries between the surface regions

This is achieved by a filter which blends the texture between neighbouring triangles of different surface regions. This results in a local blurring effect, which is less conspicuous then the blocking effect occurring without filtering.

Synthesis of texture for surface triangles not visible in any image

For that purpose a filter has been developed which uses the texture from visible triangles.

5 Results

The system has been tested with several 3D objects by using a CCD camera as well as a reflex camera in combination with a slide scanner. In Fig. 13 input images of the objects ”nana” (CCD camera), ”salt” (CCD camera) and ”statue” (reflex camera) are shown.

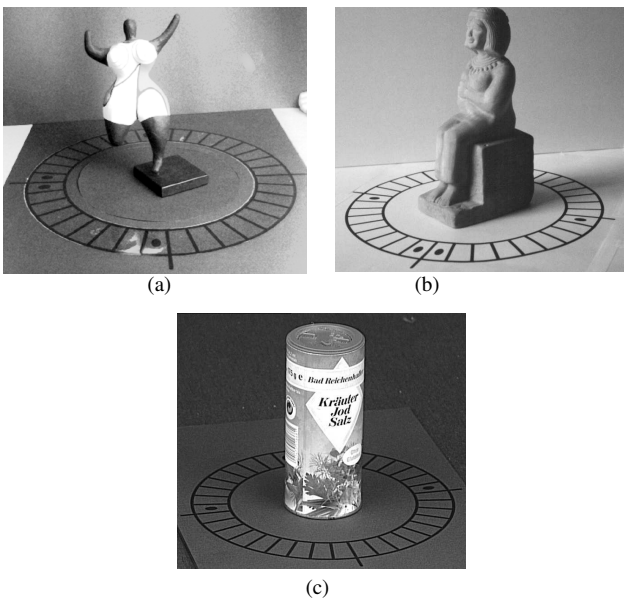


Fig. 13 Input images (a) ”nana” (b) ”statue” (c) ”salt”

In order to compare the results with those obtained by a turntable environment, the calibration pattern and the object ”salt” were put on a turntable and rotated in predefined steps of 10 °. For camera estimation only the pattern information was evaluated and the estimated rotation angle was compared with the predefined one (Fig. 14a). It turns out, that its standard deviation from the mean rotation angle of 10 ° is 0.03 °.

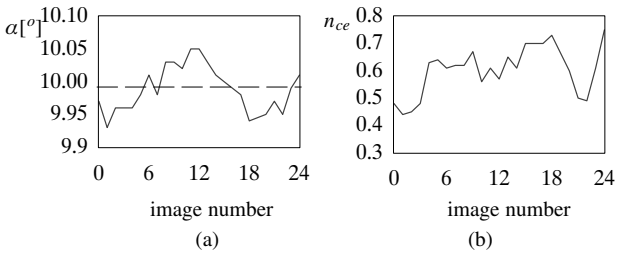


Fig. 14 (a) Estimated rotation angle α and
(b) normalized calibration error n_{ce}
for the object ”salt”

Furthermore, the normalized calibration error n_{ce} (Eq. (13)) has been evaluated for each input image and is illustrated in Fig. 14b. It shows, that $n_{ce} < 0.8$, which indicates a good calibration and is comparable to the results obtained by an advance calibration used for the turntable environment.

In Fig 16 synthezised views of some automatically generated 3D models from different positions are shown – with and without texture. It shows, that also the subjective model quality is sufficient for mulimedia applications.

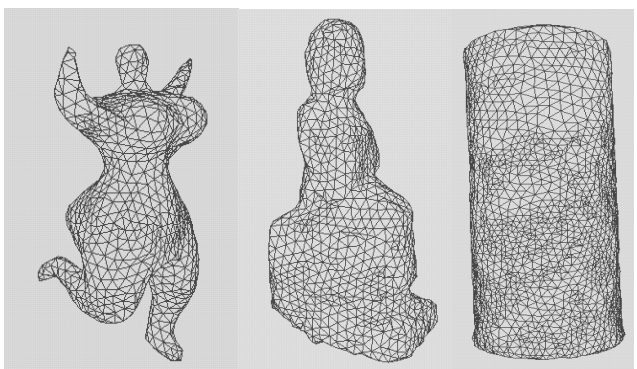


Fig. 15 Synthesized views of automatically generated
3D wireframe models
(a) 3D model ”nana”
(b) 3D model ”statue”
(c) 3D model ”salt”

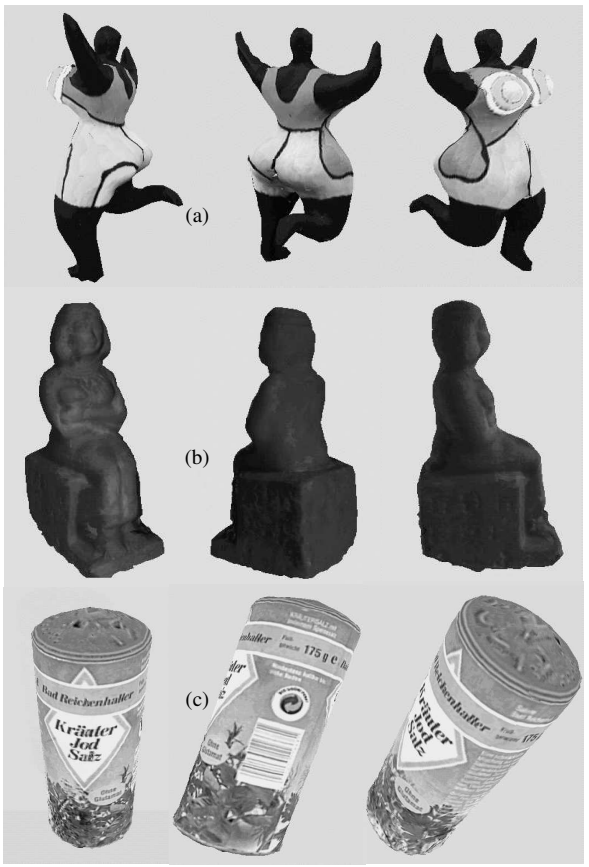


Fig. 16 Synthesized views of automatically generated
3D wireframe models with texture
(a) 3D model ”nana”
(b) 3D model ”statue”
(c) 3D model ”salt”

6 Summary

This paper presents an algorithm for the automatic reconstruction of 3D objects using a mobile monoscopic camera providing virtual 3D models for a variety of multimedia applications.

As the measurement environment should be as simple as possible, a calibration pattern is introduced which is placed below the object and allows object registration and camera calibration simultaneously for each individual input image. The pattern consists of two concentric circles connected by radial line segments with known dimensions. Together with position markers in each quadrant, a unique assignment of the camera views to a fixed world coordinate system is possible, even if parts of the pattern are occluded.

The extraction of the necessary calibration information from the input images is performed in three steps:

In a first step, object, background and calibration pattern are separated in the input images. For this purpose, a

colour segmentation technique in combination with a repeated erosion–dilatation is used.

In a second step, the 2D calibration points are estimated by intersecting the line segments and the projected circles of the calibration pattern in the image. The correspondence of the not occluded calibration points to a fixed world coordinate system is determined by evaluation of position markers.

In a third step, from the extracted 2D calibration point coordinates in the images and their corresponding 3D world coordinates, the external and internal camera parameters for each individual view are estimated by using an algorithm as proposed by Tsai.

Now a textured 3D–wireframe model is generated using a former proposed technique for the reconstruction of 3D objects from multiple calibrated camera views.

The system has been tested with different real objects. A CCD camera as well as a reflex camera in combination with a slide scanner have been used for image acquisition. It turns out, that the quality of the camera calibration is the same as obtained with the advance calibration used with turntable systems; the calculated normalized calibration error is $n_{ce} < 0.8$, which indicates a good calibration. The outlook of the resulting textured 3D models is subjectively sufficient for a variety of multimedia applications.

Compared to existing turntable systems, the this system offers a significant gain in flexibility for the user and a drastic reduction of costs for technical equipment, hence it provides new options for future 3D reconstruction techniques.

References

[1] H. Busch, "Automatic Modelling of Rigid 3D Objects Using an Analysis by Synthesis System ", *SPIE Proceedings "Visual Communication and Image Processing IV"*, 1989

[2] C. H. Chien and J. K. Aggarwal, "Identification of 3D objects from multiple silhouettes using quadtrees/octrees ", *Comp. Vision, Graphics, and Image Processing* 36, 1986, pp. 256–273.

[3] 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.

[4] W. Niem u. H. Broszio, "Mapping Texture from Multiple Camera Views onto 3D–Object Models for Computer Animation", *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging,* September 6 – 8, 1995, Santorini, Greece.

[5] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, "Numerical Recipes in C", *Cambridge University Press,* 2nd edition, Cambridge 1994.

[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] G. A. Thomas, "Global Motion Estimation for Registering Real and Synthetic Images", *Workshops in Computing– Image Processing for Broadcast and Video Production,* Springer, ISBN 3–540–199470–0, Hamburg 1994

[9] J. Weng, P. Cohen, Marc Herniou, "Camera Calibration with Distortion Models and Accuracy Evaluation", *IEEE Transactions on Pattern Analysis and Machine Intelligence* , Vol. 14. No. 10, October 1992.

[10] 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.

[11] 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.

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