

Reconstruction of Sewer Shaft Profiles from Fisheye-Lens Camera Images

Sandro Esquivel¹, Reinhard Koch¹, and Heino Rehse²

¹ Christian-Albrechts-University, Kiel, Germany

² IBAK Helmut Hunger GmbH & Co. KG, Kiel, Germany

Abstract. In this paper we propose a robust image and sensor based approach for automatic 3d model acquisition of sewer shafts from survey videos captured by a downward-looking fisheye-lens camera while lowering it into the shaft. Our approach is based on Structure from Motion adjusted to the constrained motion and scene, and involves shape recognition in order to obtain the geometry of the scene appropriately. The approach has been implemented and applied successfully to the practical stage as part of a commercial software.

1 Introduction

Automatic sewer inspection is an important application for computer vision and robotics. Remotely controlled inspection devices based on mobile robots equipped with different sensors are commonly used for this task since the observed structures are often not directly accessible for humans or access is difficult to achieve. As regular inspection of manholes and sewer shafts is required by law, this special application is interesting for commercial systems.

While different approaches to this problem exist – including solutions based on structured light, multi-frequency sonar, infrared sensors, or recently time-of-flight cameras – we rely on an approach which is based mainly on video sequences captured by a fisheye-lens camera provided with a flash light which is lowered into the manhole. Additional data acquired by a rotation sensor which is attached to the camera is used to facilitate the task. For reconstruction, we are able to assume additional constraints since the given problem of sewer shaft inspection using a hanging camera differs slightly from general sewer inspection. Figure 1 shows a commercial system using our which has been built by our industry partner IBAK Helmut Hunger GmbH & Co. KG. The video sequences used are byproducts of interactive shaft inspection.



Fig. 1. IBAK PANORAMO® SI.

Previous Work: An early idea for recovering shape and camera pose relative to the pipe axis automatically from sewer survey videos was presented in [1]. Kannala et al. [2, 3] considered an approach for automatic 3d model acquisition from video sequences captured by a calibrated fisheye-lens camera moving through a sewer pipe. They recover camera positions and scene structure by computing calibrated multi-view tensors for image sub-sequences and merging the results hierarchically, which results in a point cloud approximating the scene structure as an initial 3d model. This approach suffers from error accumulation and sensitivity to inaccurate camera calibration resulting in bent and conical pipe reconstructions which are known to be straight. Our problem formulation is slightly different since we aim to measure the shape of a shaft from a camera hanging down rather than from traveling through the sewerage.

Our Approach: The main idea of our approach is to incorporate a priori knowledge about the scene geometry and to simplify the resulting 3d model appropriately to stabilize the whole reconstruction process. Our approach computes shaft profiles at different depths by a Structure from Motion approach, classifies them as appropriate 2d shapes, and builds a 3d model by connecting shapes from subsequent cross-sections. The reconstruction is geometrically corrected using knowledge about the camera motion. Since it is designed for practical purposes, the focus of our work is on flexibility, robustness, and automation of the reconstruction process.

2 Background

2.1 Problem Specification and Setting

The setting of our work is illustrated in Fig.2: A fisheye-lens camera designed for sewerage survey is lowered vertically into a sewer shaft which is specified to be vertical with arbitrary basic shape, but often rectangular shafts or shafts with elliptical profile. Images are captured in fixed translation intervals which can be measured accurately from the feed of the conducting cable (in our case, the camera moves up to 35 cm/s, but a flash ensures sharp images every 5 cm). Additionally, an inertial sensor is mounted to the camera which measures roll rotation around the viewing axis for each image to compensate this rotation later in the images. While it is assumed that the camera is looking approximately along the axis of the shaft, the exact position of the camera is unknown. The camera might also oscillate around the cable axis. The task is to classify and measure the cross-sectional shape of the shaft at different depths robustly and obtain an approximate 3d model of the shaft by appropriately merging profiles from subsequent cross-sections.

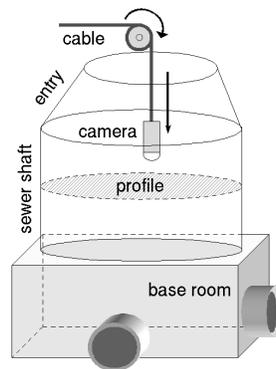


Fig. 2. Setup for inspection.

Figure 3 shows typical input images captured by the fisheye-lens camera during lowering it into a sewer shaft through the manhole. Apparently, the task of visual reconstruction is not trivial: Illumination and visibility decrease rapidly towards the center of the image, the hanging camera is rotating significantly around its view axis, there are reflections especially on fronto-parallel parts of the shaft surface and obscuring structures such as stairs and branching pipes, and vision is very poor in larger rooms where the camera is located off-center.

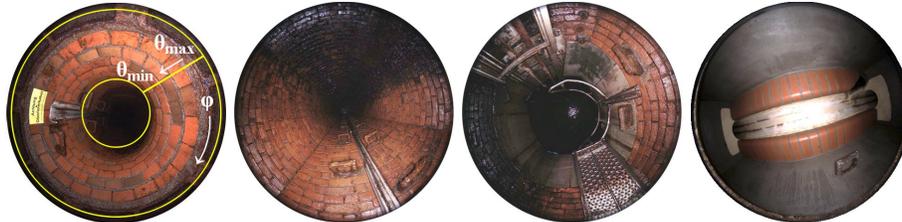


Fig. 3. Input images captured by a hanging fisheye-lens camera during lowering.

2.2 3D Reconstruction Using Structure from Motion

There are several Structure from Motion (SfM) approaches using spherical cameras such as the ones described in [6], [2, 3] and [7]: Chang and Hebert [6] describe a SfM approach for general scenes using cameras with wide field of view such as fisheye-lens cameras and analyze its uncertainty. They show that SfM performs better in certain situations (e.g. sparse scene structure, motion along the viewing direction of the camera) which apply to the given setting.

Kannala et al. [2, 3] compute sparse structure of sewer pipes from feature point correspondences for image triplets by estimating the trifocal tensor, and merge local reconstructions by hierarchical bundle adjustment.

We use an approach by Bartczak et al. [7]. They describe SfM from a dense sequence of spherical images of a rigid scene with no a priori knowledge about camera motion which avoids heuristics by heavily using information about measurement uncertainties and error propagation. The reconstruction process is separated into a bootstrapping and a tracking stage based on 2d-2d correspondences between subsequent images. During bootstrapping, an initial sparse scene structure and camera position is computed from the essential matrix estimated from 2d-2d correspondences between the first image pair. During tracking, subsequent camera poses are computed from 2d-3d correspondences while new scene structure is estimated from 2d-2d correspondences and existing 3d structure is updated using further 2d measurements. Bundle adjustment is used after bootstrapping to ensure a good initialization.

All mentioned approaches rely on multi-frame feature point correspondences (*trails*) obtained using the well-known KLT feature tracker [8]. Because the input video sequence is rather sparse in our setting, feature prediction between subsequent images is inevitable, and bundle adjustment as a final step is not useful since feature trails are in average very short (3–5 images).

Note that without knowledge about the distance between the first two camera positions, 3d structure can only be estimated up to an unknown scale. In our setting, we can overcome this ambiguity since the translation amount of the camera between subsequent image captures is known to be approximately 5 cm.

In [3], Kannala et al. derive the course of the pipe from the camera path. Nevertheless there is no reasonable distinction between error accumulation on the camera motion and factual curvature of the pipe. This can result in bent reconstructions of essentially straight pipes. Since the feature point correspondences from which camera motion is estimated can only be obtained for very few subsequent images, drift of the reconstructed camera path is very likely as has been observed in our experiments. In contrast to Kannala’s approach, in our setting the camera path is known to be oscillating around a straight line – given by the vector of gravity – which will allow us to correct the reconstruction.

3 Our Approach

Our algorithm is composed from the following steps which will be explained in detail in the following sections:

1. Reconstruction of 3d points on the shaft surface
 - (a) Cylinder-mapping of input camera images, removing roll rotation using the input of an additional rotation sensor (pre-processing step).
 - (b) Structure from Motion as described in [7] using cylinder-mapped images with problem-specific feature prediction (tracking phase).
 - (c) Correction of reconstructed geometry and camera motion using a priori knowledge about motion and scene geometry (post-processing step).
2. Contour shape classification and shape fitting in cross-sections of the shaft, and construction of a simple 3d geometry by connecting contours from subsequent cross-sections and optional 3d shape fitting.

Prior to application, common calibration techniques for fisheye-lens cameras [5] are used in order to estimate the intrinsic parameters and radial distortion of the camera. While the camera used is almost distortion-free, note that we allow the focal length of the camera calibration to have an error of up to several percent for sake of robustness.

3.1 Reconstruction of 3D Points

Reconstruction of 3d points relies on the following assumptions which are supposed to hold for the given problem setting:

- The viewing direction of the camera is mainly along the shaft axis with up to 5–10° pan/tilt rotation and almost no roll rotation.
- The average motion vector of the camera coincides with the shaft axis.
- Within shaft sections, the camera-local shaft profile changes only slightly between subsequent images due to small transversal motion and pan/tilt rotation of the camera, or continuous profile changes (e.g. conical sections).
- Abrupt changes of the entire camera-local shaft profile indicate geometry changes of the shaft (e.g. at the junction of shaft and base room).

Cylinder-mapping of camera images. Existing approaches for image-based sewer reconstruction detect and track feature points directly in the spherical camera images resp. apply local perspective undistortion first [2, 3]. In our work we determined that “unwinding” the image according to spherical coordinates as seen in Fig.4 (left) – approximating an image of the unrolled shaft surface – facilitates the feature tracking process as long as the camera’s viewing direction is approximately parallel to the shaft axis. We account for a ring-shaped part of the camera’s field of view corresponding to the viewing range between zenith angles $\theta_{\max} \geq \theta \geq \theta_{\min}$ (here: $\theta_{\max} := 85^\circ$, $\theta_{\min} := 45^\circ$) which holds the most usable visual information (compare Fig.3 (left) and Fig.4 (left)).

Rotation ϕ_{RS} around the camera’s z -axis measured by an inertial sensor is used during mapping to compensate roll rotation of the camera. This is easy since roll rotation results simply in a vertical shift of the cylinder image.

Tracking points in cylinder-mapped images. For feature detection and tracking in the cylinder-mapped images we use an implementation of the KLT feature tracker [8]. Since the displacement of image points in subsequent images is large, even for multi-resolution tracking, either feature prediction or region search is necessary. Tracking points in cylinder-mapped images is based on the fact that feature points move mainly along image rows, and disparity deviation is distinctive for points in different image columns but only small within the same row. Figure 4 illustrates the disparity between corresponding cylinder image points for an exemplary motion along the cylinder axis with off-center position, small transversal motion and 5° pan/tilt rotation. The displacement vectors of corresponding image points are basically horizontal, average disparity varies strongly with respect to vertical image position but only little with respect to horizontal image position. The latter curve depends on the cross-sectional shape of the cylinder and the excentricity of the camera. The former curve is affected by roll rotation between images which is minimal since it is compensated during mapping.

As a consequence, the following feature tracking method is proposed:

Row Scan (Init): Without any knowledge about the shaft’s diameter and profile shape, feature point correspondences are generated by scanning the image row and computing a similarity measure between feature points detected in the previous and current images. The vertical tolerance is defined by the maximal valid distance to the shaft surface (here: 300 cm) and maximal pan/tilt rotation (here: 10°). This method is used for initialization and for re-initialization when tracking fails during SfM (e.g. due to noticeable shaft geometry change).

Row Track: After initialization, the average row-wise disparity of the last images is used to predict feature point positions within the current image.

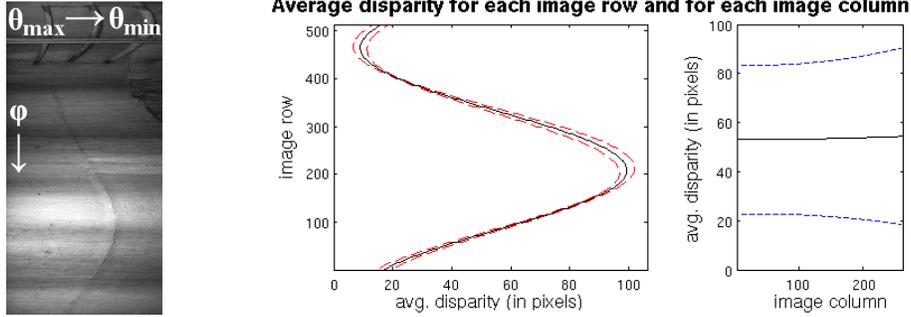


Fig. 4. Average column/row-wise disparities for corresponding cylinder image points.

3.2 Geometric Correction

As described above, in SfM the camera motion estimation tends to drift which results in globally erroneous reconstructions as shown in Fig.5 (left):

The shaft will be bent and its diameter will narrow resp. widen over time. The latter effect occurs prominently as a systematical error when the focal length of the fisheye-lens camera has not been calibrated correctly before application. Common solutions to the problem of error accumulation, such as multi-frame SfM, depend on tracking feature points through a large number of frames and thus can not be applied here. On the other hand, the reconstruction is nevertheless locally correct since inter-frame errors are small. In the following, a simple approach is described how to correct the reconstructed camera poses. Once this is done, structure is corrected locally with respect to the camera by which it was originally seen, resulting in a globally consistent reconstruction.

In our setting, the camera is hanging into the shaft by its own weight on a static cable. Image acquisition is triggered at certain equidistant amounts of feed (5 cm). Hence for camera pose correction we can assume (a) that the average camera path approximates the vector of gravity, and (b) that the distance between subsequent camera positions along the vector of gravity is fixed and known.

Geometric correction (GC) is accomplished hence as follows:

1. First, approximate the mean “drifted” camera motion by fitting a polynomial curve $\mathbf{p}(t) := (p_x(t), p_y(t), t)$ to all camera positions $\mathbf{C}_0, \dots, \mathbf{C}_N$.
2. Correct camera positions $\mathbf{C}_0, \dots, \mathbf{C}_N$ by “unbending” the mean camera motion so that it is mapped to the world z -axis (i.e. the vector of gravity).
3. Correct camera positions further by rescaling camera motion locally so that inter-camera distance is equalized to 5 cm each.

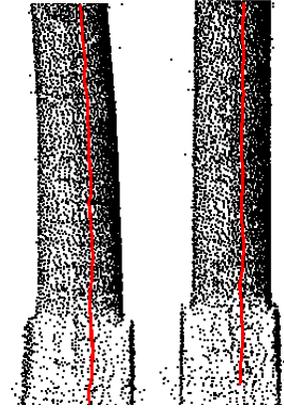


Fig. 5. 3d points before (left) and after geometric correction (right).

4. Correct pan/tilt rotation of each camera such that the local “drifted” gravity vector $\nabla \mathbf{p}(t)$ becomes parallel to the world z -axis. The estimation of camera roll rotation is assumed to be drift-free.
5. Finally, the positions of all 3d points \mathbf{X}_j are updated with respect to the new pose of the first camera \mathbf{C}_i they were visible in.

Note that geometric correction transfers all 3d points and camera poses into a common coordinate frame which is registered to the (ideal) shaft geometry – i.e. z -axis is parallel to shaft axis and x/y -plane is parallel to cross-sections.

3.3 Shape Classification and Estimation

The goal of the next stage is to estimate the average cross-sectional shape of the shaft at M evenly distributed depths h_0, \dots, h_{M-1} (here: $h_i := i \cdot 5$ cm each). First, the 3d points are partitioned into M slices $\mathcal{S}_0, \dots, \mathcal{S}_{M-1}$ where each slice \mathcal{S}_i consists of the 2d projections (x_j, y_j) of all 3d points (x_j, y_j, z_j) with $z_j \in [h_i, h_{i+1}[$ onto the x/y -plane. Common 2d shape fitting methods – such as [10] for ellipses, [11] for rectangles or [12] for closed spline curves – are used to obtain shape estimates for each slice \mathcal{S}_i . To enable robustness against 3d points that do not lay on the shaft surface or result from incorrect triangulation, a RANSAC approach is used in combination with the shape fitting methods.

The classification is done by fitting an instance of each shape class to all 2d points in slice \mathcal{S}_i robustly, evaluating a quality score for each shape which is based on the average geometric distance of all inliers and weighted such that shape class changes are punished, and selecting the shape with the highest resulting score.

3.4 3D Model Creation from Cross-Sections

In general, the structure of sewer shafts can be modelled as a sequence of straight segments of extrusion-like geometries (i.e. generalized cylinders) without branches. Hence the 3d surface can be constructed simply by connecting subsequent contours of the same shape class with significantly small difference in shape parameters and interpolating linearly between cross-sections. The reconstructed model can be further simplified by fitting special extrusion surfaces (e.g. cylinders for elliptical cross-sectional shape, cuboids for rectangular shape) to 3d points within each segment (see e.g. [9]).

4 Experiments and Results

4.1 Evaluation with Real Labelled Data

To evaluate the performance of our approach with real data, our industry partner has provided us with a number of video sequences that has been captured from different sewer shafts (44 sequences from 36 different shafts). The observed shafts show a great variety of depth, diameter and shape. First, we identified the shape and diameter for 60 subsequences (“reference sections”) manually using

previous knowledge about the parts the shafts are made up from. We applied our algorithm to each video sequence and evaluated for each reference section if the correct shape class was identified and measured the average estimation error for each correctly classified cross-section by comparing it with the manual reference data. For elliptical cross-sections the average diameter error is regarded, for rectangular cross-sections the average lateral length error. The results are shown in Fig.6. For each reference section, the average diameter estimation error and the standard deviation of the errors are shown. In order to evaluate the performance of the proposed geometric correction, we applied our implementation once with and again without geometric correction. Note that the last 3 sequences have in fact pulvinate rectangular shape. Our approach failed for 3 reference sequences that are not shown in Fig.6 which have pentagonal shape.

Apparently, the average relative error is ca. 1–2% which corresponds to an absolute error of ca. 2 cm in diameter resp. lateral length. Since the reference data is idealized and does not pay attention to possible local deformations of the shafts, the comparison has to be interpreted rather as a verification of our approach than as an exact evaluation of accuracy. Note also that using geometric correction the estimated cross-sectional diameters vary less than without geometric correction and the overall accuracy improves significantly. Geometric correction is also capable of compensating deviating scale errors resulting from inaccurate focal length calibration of the camera as shown in Fig.7 for one shaft.

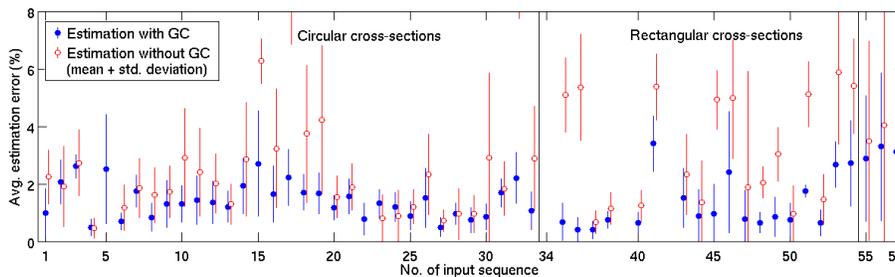


Fig. 6. Results of our algorithm with and without geometric correction (GC) for 57 out of 60 shaft segments with fixed and approximately known diameter.

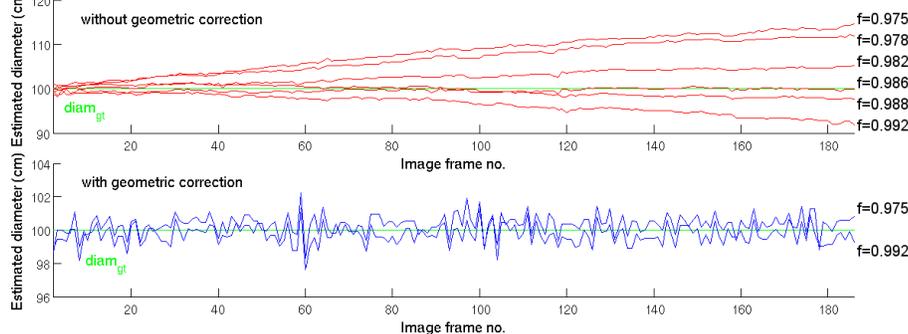


Fig. 7. Estimation results for shaft sequence no. 8 with varying focal length f without GC (top) and with GC (bottom). Note the different scales of the graphs.

4.2 Practical Issues: Robustness and Runtime

Contour-based model generation with GC failed only for 3 total sequences and 3 subsequences out of the test set of 44 sequences: 3 examples show a shaft with non-standard pentagonal geometry which is not supported up to now. The other reconstructions failed due to very poor vision and strong reflections, both concerned shafts consist mainly of the base room (5–6 images) following a very short shaft. Nevertheless, SfM and GC succeeded for all sequences but resulted in very sparse point clouds for the latter 2 shafts. Without GC, model generation failed for 2 more shafts where the GC approach succeeded. Test applications done by our industrial partner with more than 160 shafts yielded similar results: Contour-based reconstruction failed for 4 shafts due to illumination/visibility problems while point-based reconstruction succeeded always with plausible results.

The total runtime of our implementation is basically linear in the number of input images. Repeated tests with all provided test sequences yielded an average factor of 0.45 ± 0.07 sec per image on a PC with 2.66 GHz CPU and 4 GB RAM. By further optimization the runtime is expected to be improved significantly, e.g. by utilizing the GPU for tasks such as computing the cylinder-mapping which consumes ca. 30% of the total runtime at the moment. Nevertheless, the runtime is already acceptable for post-processing of image sequences of up to 500 frames (i.e. 10 m of shaft) in clearly less than 5 min.

4.3 Resulting 3D Models

Using our algorithm we were able to build simplified 3d models of the surveyed shafts. Figure 8 shows the reconstruction results (i.e. the corrected 3d points resulting from the SfM, and the original and simplified wire frame mesh of the identified contours) for one of the shafts which consists of a conical part below the man-hole, a cylindrical main part, and a cubic base room.

The example illustrates that our SfM approach is capable of recovering even fine structure reliably such as the stairs or the channel at the ground of the shaft (Fig.8, left) while the contour classification is robust enough to regard such structures as outliers with respect to the basic shape (Fig.8, right). Although we build only wire frame models from the resulting geometry, standard texture mapping techniques could be used to reconstruct a fully textured 3d model from the video sequences and the reconstructed geometry which allows to navigate virtually through the shaft under survey and perform measurements.

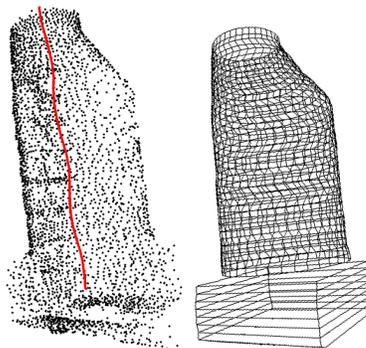


Fig. 8. Corrected 3d points with camera path and wire frame model.

5 Conclusions

We have proposed a robust practical approach for automatic shape measuring and 3d reconstruction of sewer shafts using a fisheye-lens camera provided with an inertial sensor unit. Our approach overcomes the problems determined by similar works considering the problem of building 3d models for sewerage, such as bent or conical reconstruction and restriction to elliptical pipes [2, 3]. It can easily be extended to support other shaft shapes than ellipses, rectangles, and free-form curves on demand, e.g. ovoid or polygonal shapes. An implementation has been applied successfully to the practical stage in cooperation with our industrial partner IBAK as part of the software for the widely used PANORAMO[®] SI system (see Fig.1). Practical test applications, i.a. done by the Göttinger Entsorgungsbetriebe [13], have shown that our approach is robust and useful.

Future Work: Up to now our approach is performed as a post-processing step. We are planning to merge all parts of our approach into an online process. By building an approximate 3d model during tracking, a more elaborate feature prediction can be considered by projecting feature points onto the estimated scene surface – approaching an on-the-fly Analysis by Synthesis technique.

References

1. D. Cooper, T. P. Pridmore, N. Taylor: Towards the Recovery of Extrinsic Camera Parameters from Video Records of Sewer Surveys. *Machine Vision and Applications* 11, pp. 53–63, 1998.
2. J. Kannala: Measuring the Shape of Sewer Pipes from Video. Master thesis, Helsinki University of Technology, Helsinki, 2004.
3. J. Kannala, S. S. Brandt, J. Heikkilä: Measuring and Modelling Sewer Pipes from Video. *Machine Vision and Applications* 19(2), pp. 73–83, 2008.
4. Z. Zhang: Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. *Proc. ICCV*, pp. 666–673, 1999.
5. D. Scaramuzza, A. Martinelli, R. Siegwart: A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion. *Proc. ICVS*, p. 45, 2006.
6. P. Chang, M. Hebert: Omni-Directional Structure from Motion. *Proc. IEEE Workshop on Omnidirectional Vision*, p. 127, 2000.
7. B. Bartczak, K. Köser, F. Woelk, R. Koch: Extraction of 3D Freeform Surfaces as Visual Landmarks for Real-Time Tracking. *Journal of Real-Time Image Processing* 2(2–3), pp. 81–101, 11/2007.
8. C. Tomasi, T. Kanade: Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, 04/1991.
9. V. Pratt: Direct Least Squares Fitting of Algebraic Surfaces. *Computer Graphics* 21 (4), pp. 145–152, 1987.
10. A. Fitzgibbon, M. Pilu, R. Fisher: Direct Least Squares Fitting of Ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(5), pp. 476–480, 1999.
11. D. Chauduri, A. Samal: A Simple Method for Fitting of Boundary Rectangles to Closed Regions. *Pattern Recognition* 40(7), pp. 1981–1989, 2007.
12. P. Dierckx: *Curve and Surface Fitting with Splines*. Oxford University Press, 1993.
13. B. Burger, M. Fiedler, J. Gellrich, H.-P. Reuter: Schachtinspektion in neuer Qualität. *Bauwirtschaftliche Information UmweltBau* Nr. 1, 02/2009.