3D Reconstruction of Sewer Shafts from Video

Sandro Esquivel[1], Reinhard Koch[1], Heino Rehse[2]

[1] Christian-Albrechts-University of Kiel
Hermann-Rodewald-Str. 3, 24118 Kiel, Germany

[2] IBAK Helmut Hunger GmbH & Co. KG
Wehdenweg 122, 24148 Kiel, Germany

esquivel@mip.informatik.uni-kiel.de

Abstract: In this paper we propose a robust approach for automatic 3d model acquisition of sewer shafts from survey videos. Images are captured by a specific camera setup which is composed of a downward-looking fisheye-lens camera while lowering it into the shaft. Additionally, an inertial sensor measures rotation around the viewing axis. Our approach is based on Structure from Motion adjusted to the constrained motion and scene geometry, and measures the profile of the shaft using robust 2d shape recognition techniques. Global bundle adjustment is avoided by applying a simple and fast geometric correction of the computed 3d reconstruction which is also capable of handling inaccuracies of the intrinsic camera calibration parameters. An implementation of our method has been evaluated extensively with real data. Furthermore, we have proposed modifications of our so far off-line implementation to approach real-time reconstruction which can be applied during on-site inspection.

1 Introduction

Automatic 3d reconstruction from video and sensor data is a very important topic of research in photogrammetry and computer vision, and has been largely studied. Recent systems are capable of real-time reconstruction by executing time-consuming parts in parallel on the GPU. While different approaches exist - including systems using active sensors such as laser scanners, sonar, or recently time-of-flight cameras - purely visual methods are still interesting to enhance existing systems where video is generated as a byproduct of manual inspection, and to reduce production costs.

An interesting application for 3d reconstruction is the support of sewer and sewer shaft inspection systems. Remotely controlled inspection devices such as mobile robots equipped with cameras and sensors are commonly used for this task since the concerning structures are often not directly accessible for humans or access is difficult to achieve. Since regular inspection of manholes and sewer shafts is required by law, there is a demand for commercial systems for this special application. Conventional sewer inspection systems are remote-controlled and capture visual data which is analyzed by experts. In order to facilitate the surveillance process, commercial sewer inspection systems are required to measure the 3d geometry of the scene automatically which can also be used for later visualization. In this paper we present a robust approach for automatic 3d reconstruction of sewer shafts using a specific camera setup which is built and delivered by our industry partner IBAK Helmut Hunger GmbH & Co. KG, and is in use for sewer shaft inspection at several locations.

1.1 Previous and Related Work

3d scene reconstruction from video is commonly solved by Structure from Motion (SfM) which simultaneously estimates the camera motion and sparse scene structure from corresponding points in subsequent camera images [HZ03]. Recently there has been a lot of work on porting computer vision algorithms to the GPU, resulting in near real-time SfM implementations. Accurate reconstruction demands though for a global optimization of the scene geometry and camera parameters as a final step, most often by bundle adjustment which is very time-consuming [TMH99].

While the literature about SfM in general is abundant, there has also been some previous work focusing explicitly on the 3d reconstruction of sewers. An early idea for recovering shape and camera pose relative to the pipe axis from sewer survey videos was presented in [CPT98]. Kannala et al. considered an approach for automatic 3d model acquisition from video sequences captured by a calibrated fisheye-lens camera moving through a sewer pipe [KBH08]. 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 yet 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 hanging camera as described in the next section.

1.2 Problem Specification and Setting

A sketch of the sewer inspection setup as well as a picture of the commercial system provided by our industry partner is shown in fig. 1: A fisheye-lens camera 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. Color images of size 1040×1040 pel are captured in fixed translation intervals which can be measured accurately from the feed of the conducting cable. In the given setting, the camera moves up to 35 cm/s and captures images with 7 Hz every 5 cm. A flash ensures sharp images within in the shaft (see fig. 2). Additional, 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 merging profiles from subsequent cross-sections appropriately. We approach this problem by designing a robust SfM pipeline which is presented shortly in the following.

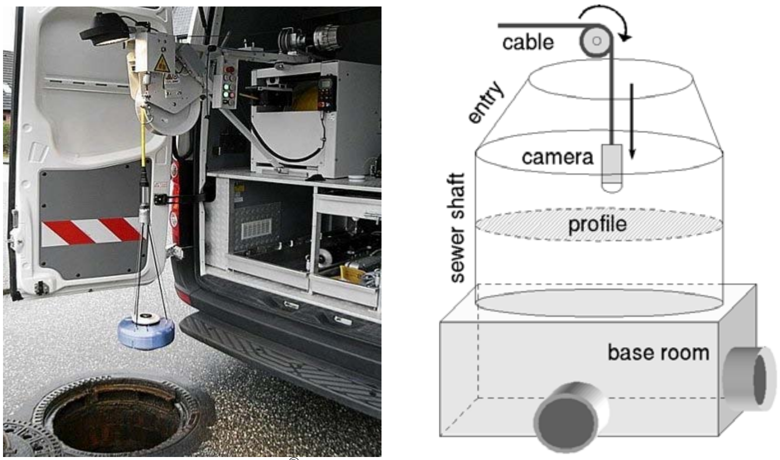

Figure 1: IBAK PANORAMO® SI and schematic setup for sewer shaft inspection

1.3 Our Approach

The main goal of our approach is to exploit knowledge about scene geometry and camera motion to constrain the reconstruction process in order to stabilize the algorithm. Computed camera poses and sparse scene structure are used to measure shaft profiles at different depths, classify them as appropriate 2d shapes, and build a 3d model by connecting shapes from subsequent cross-sections which can be visualized (see fig. 3) or used for further manual measurements. A crucial contribution is a novel method for global optimization of the resulting geometry using a computationally very efficient method based on knowledge about the camera trajectory rather than using classical bundle adjustment. We further explore possibilities to parallelize parts of the algorithm and execute them on the GPU such that near real-time 3d reconstruction on site will become feasible.

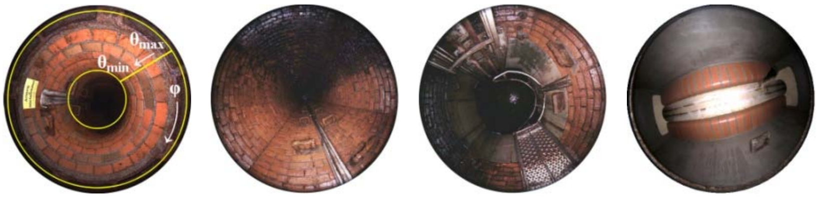

Figure 2: Input images captured by a fisheye-lens camera lowered into the shaft

2 Our Approach to 3D Shaft Reconstruction

The main pipeline of our approach is composed of the following steps which will be discussed in detail in the following sections (see fig. 4):

- **Preprocessing:** Cylinder-mapping of input images, registering images using the input of an additional rotation sensor
- **Structure from Motion:** Detection and tracking image points with adaptive prediction, simultaneous reconstruction of sparse scene geometry and camera motion
- **Post-processing:** Global optimization of the scene geometry and camera path
- **3d model creation:** Shape classification of cross-sections, creation of 3d model

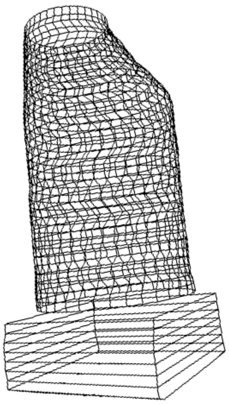

Figure 3: Resulting 3d model of sewer shaft

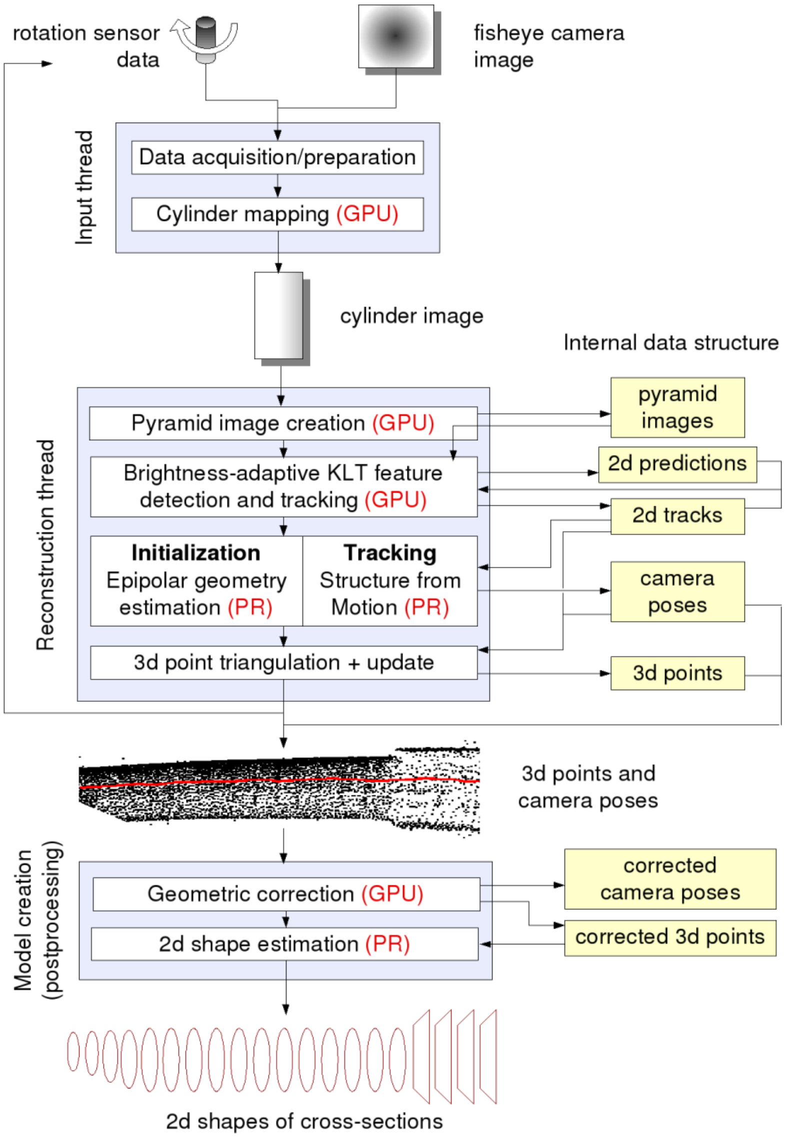
rotation sensor data
fisheye camera image
Input thread
Data acquisition/preparation
Cylinder mapping (GPU)
cylinder image
Internal data structure
Reconstruction thread
Pyramid image creation (GPU)
Brightness-adaptive KLT feature detection and tracking (GPU)
Initialization
Epipolar geometry estimation (PR)
Tracking
Structure from Motion (PR)
3d point triangulation + update
pyramid images
2d predictions
2d tracks
camera poses
3d points
3d points and camera poses
Model creation (postprocessing)
Geometric correction (GPU)
2d shape estimation (PR)
corrected camera poses
corrected 3d points
2d shapes of cross-sections

Figure 4: Overview of the main processing pipeline of our algorithm. Input images and rotation sensor data are preprocessed by an input thread. Computation of sparse scene structure and camera motion is done by a concurrent thread. Finally, the reconstruction is globally optimized and cross-sectional shapes are identified. Potential computational speedup is denoted by GPU and PR (see sec 2.6).

2.1 Camera Model

Since we use fisheye-lens cameras with minimal radial distortion, the image formation process can be modeled by a simple equiangular spherical mapping of image points (u,v) to 3d rays (φ,θ) within the camera coordinate frame. The mapping depends only on the principal point (p_u,p_v) and radius r of the 90 degree circle within the camera image. We assume the cameras to be calibrated but aim at high tolerance against calibration inaccuracies.

2.2 Image Preprocessing

Although existing SfM approaches detect and match prominent image points in the fisheye images directly, we showed in [EKR09] that for the specific scene geometry, point tracking benefits significantly from mapping the ring-shaped region of interest in the images to cylinder coordinates first, approximating an image of the unwrapped local shaft surface (see fig. 5). Rotation around the viewing axis is compensated using the input of the rotation sensor.

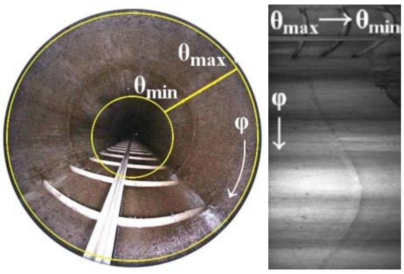

Figure 5: Fisheye image with region of interest and cylinder-mapped image

2.3 Reconstruction of Camera Poses and 3D Points

Structure from Motion (SfM) computes the camera positions and sparse scene geometry from image point correspondences between subsequent images. Once correspondences have been computed, the initial step is to exploit the epipolar geometry between the first image pair to compute the relative pose up to scale and triangulate 3d points. Afterwards, the pose is computed by tracking 2d/3d correspondences and triangulating new 3d points from 2d/2d correspondences on the fly. 3d points that are visible in multiple images are tracked further, and their positions are updated with the most recent camera pose via an Extended Kalman filter [JU97] in order to increase accuracy. In case that the SfM procedure fails due to bad visibility conditions, abruptly changing scene geometry or computational errors, the algorithm is reinitialized from 2d/2d correspondences. Image point tracking is done in the rotation-compensated cylinder images using the KLT feature tracker [TK91] modified for brightness invariance. Since the distance between projections of the same 3d point in subsequent images is rather large (up to 50 pel), 2d positions must be predicted appropriately. Figure 6 shows the average offset between corresponding 2d points in subsequent images. Apparently, optical flow in the cylinder-mapped images is mainly restrained to the x-axis. We extend feature tracking by an adaptive row-wise prediction which is initialized by a row-scan for the best match for each image point.

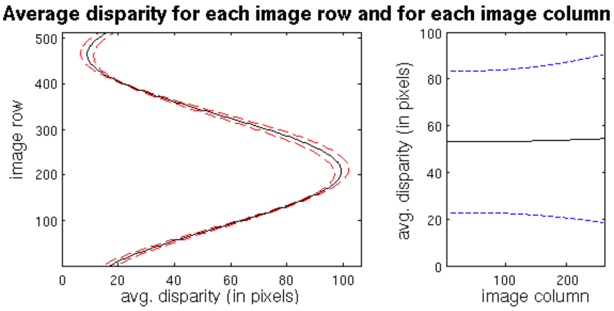

Figure 6: Mean and standard deviation of distance between 2d points in subsequent cylinder-mapped images of size 256×512 depending on image row and column

2.4 Global Optimization of 3D Reconstruction

While the final step in SfM is most commonly a global optimization of all 2d/3d correspondences using the computationally very expensive bundle adjustment method, we developed a very simple global correction of the 3d reconstruction and camera motion based on knowledge about the motion. Since the camera is lowered into the shaft hanging on a cable without lateral forces, the average camera path is known to approximate the vector of gravity. Additionally, the distance between subsequent camera positions along the vector of gravity is approximately known from the cable feed (ca. 5 cm/frame).

Due to camera calibration inaccuracies and error accumulation during the SfM procedure, the resulting average camera trajectory appears to be a curve with increasing or decreasing velocity. Hence the reconstructed shaft geometry is bent and bulged which is also noted in related work [KBH08]. We use both camera motion constraints to compute a non-linear mapping from the estimated average camera trajectory to the z-axis with even spacing between camera centers along the z-axis. Cameras and 3d points are then geometrically corrected by this mapping as shown in fig. 7. For the details of the correction algorithm see [EKR09].

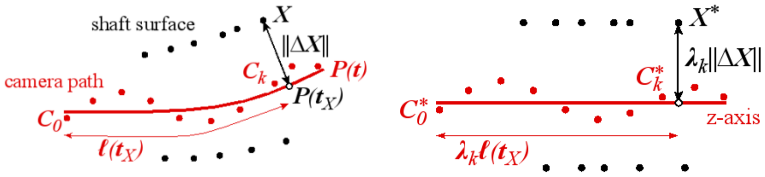

Figure 7: Camera positions C_i, average camera path P(t), and reconstructed 3d points X before (left) and after global geometric optimization (right). P(t_X) denotes the orthogonal projection of X onto P(t) which is corrected by a local scale λ_k, and repositioned at the z-axis with respect to its previous distance to P(t) (see [EKR09]).

2.5 Classification of Cross-Sectional Shapes

After 3d reconstruction and global optimization of 3d points and cameras poses, the cross-sectional shapes of the shaft at the camera locations is estimated by fitting instances of different 2d shape classes to the ortho-projection of 3d points within ±2.5cm range with respect to the z-axis robustly using a RANSAC approach [FB81]. The score for shape selection is computed from the number of RANSAC inliers and a penalizing term for the change of the shape class with respect to the previous cross-section.

2.6 Real-Time 3D Reconstruction

Since our original approach was intended for offline application, run-time was no crucial issue for our implementation. In [EKR10] we performed a time-budget analysis, identified time-expensive steps of our algorithm and proposed modifications to approach a real-time method which can be used online within approximately the same frame rate as image capturing (7 Hz). Feature detection and matching appeared to be the major time-consuming subroutines of our algorithm. To approach real-time performance, we integrated the GPU KLT tracker implementation by Zach et al. [ZGF08] which performs on 512×1024 pel images with >50Hz. Robust estimation of camera poses and cross-sectional shapes is another bottleneck in our pipeline since many iterations must be evaluated when the data has significant outlier rates as in our case. The traditional RANSAC scheme [FB81] can be replaced by similar robust estimation schemes designed for real-time application such as the preemptive RANSAC scheme proposed by Nistér in [Ni05] which is expected to provide a speedup of about 50%. Nevertheless, global optimization as described in sec. 2.4 cannot be applied as a single post-processing step in an online approach but must be repeated at certain intervals during reconstruction (e. g. every 10 frames).

3. Experiments and Results

An implementation of our approach in C++ has been tested extensively with a set of 44 real video sequences provided by our industrial partner using the camera setup shown in fig. 1. The observed shafts show a great variety in depth, diameter and shape. To evaluate the results, the cross-sectional shapes of the shafts were manually measured and labeled. In the following, we will present the results for shape measuring accuracy, robustness against calibration errors, and the runtime of the offline and online approaches.

3.1 Evaluation with Real Data

We applied our implementation to 57 subsequences of the test set consisting of 28 up to >300 images, and compared the results from shape classification and measuring with the manually measured ground truth. The results are shown in fig. 8. For each section, the average diameter estimation error and the standard deviation is shown. The average relative error is ca. 1-2% which corresponds to an absolute error of ca. 2 cm in diameter resp. lateral length. Note that the last 3 sequences have in fact pulvinate rectangular shape. Our approach failed for a total of 5 reference sequences. 3 sequences that are not shown in fig. 8 refer to shafts with pentagonal shape. The other sequences showed base rooms with very poor vision and strong reflections on the ground. Since the reference data 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 geometric correction stabilizes the estimation and contributes to the accuracy of the measurement significantly.

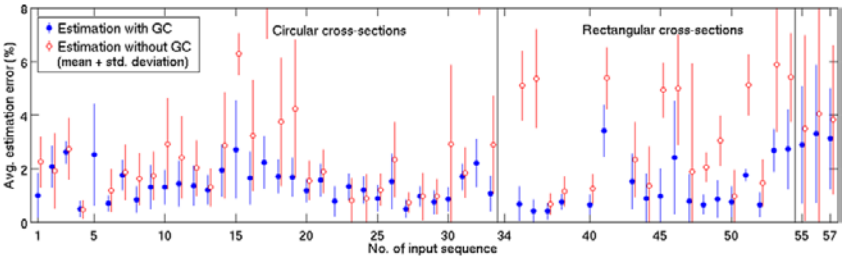

Figure 8: Accuracy of shaft diameter estimation for 57 test sequences

3.2 Robustness of our Approach

As described in sec. 2.4, the reconstruction suffers significantly from inaccuracies of the intrinsic camera calibration resulting in systematic errors. We showed that the proposed global optimization method is capable of compensating such effects largely as shown in fig. 9 while having very low computational demands (see table 1).

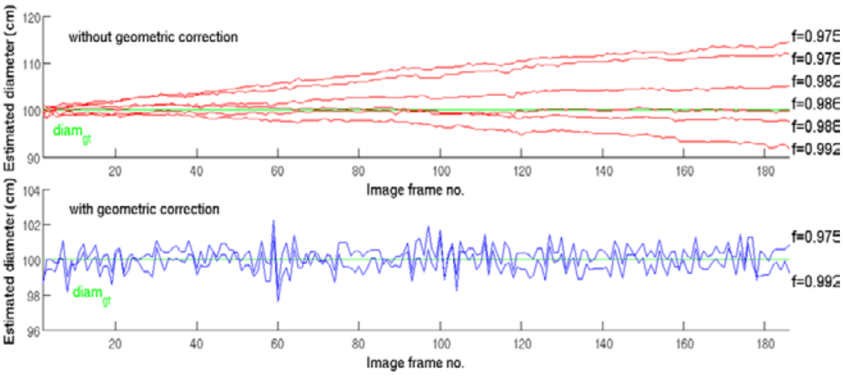

Figure 9: Results for shaft diameter estimation in sequence no. 8 with varying focal length estimates f without (top) and with global geometric correction (bottom). The deviance over time is reduced significantly (note the different scales of the graphs).

3.3 Runtime Comparison of Offline and Real-Time Approach

We measured the runtimes of our implementation for 49 test sequences consisting of 28-390 images each (ca. 4800 images in total) and compared them with the expected times of the modified online approach proposed in sec. 2.6 . Table 1 lists the average times consumed by different subroutines of both approaches. While the offline algorithm shows an average frame rate of 2.4 Hz, we approach the required limit of 7 Hz given by the image capture rate of the camera setup in the modified algorithm.

4. Conclusion

We have proposed a robust practical approach for automatic 3d reconstruction of sewer shafts using a fisheye-lens camera supported by a rotation sensor. Our approach overcomes the problems determined by similar works considering the problem of building 3d models for sewerage, such as bent or conical reconstructions and restriction to elliptical profiles

Time/frame for step	Offline approach	Online approach
Image preprocessing	120.4 ± 14.8 ms	47.5 ± 8.0 ms
KLT tracking (init.)	431.8 ± 181.5 ms	46.1 ± 1.5 ms
KLT tracking	152.1 ± 43.9 ms	15.1 ± 1.1 ms
Pose estimation (init.)	196.4 ± 69.1 ms	< 100 ms
Pose estimation	55.9 ± 26.5 ms	< 30 ms
3d triangulation/update	45.4 ± 18.6 ms	45.4 ± 18.6 ms
Geometric correction	1.3 ± 0.4 ms	6.8 ± 1.6 ms
Shape estimation	8.0 ± 2.2 ms	< 5 ms
Total runtime	**410.1 ± 168.8 ms**	**< 140 ms**

Table 1: Runtime comparison between offline and online approach (note that the time-expensive initialization steps are performed only 2-3 times per sequence)

[KBH08]. An implementation of our approach is used successfully in practical applications as part of a commercial software for the widely used PANORAMO® SI system (see fig. 1) delivered by our industry partner IBAK Helmut Hunger GmbH & Co. KG. Our approach has proved useful in practical evaluations, for example done by the Göttinger Entsorgungsbetriebe [BFG09]. Furthermore, we have analyzed the runtime of our implementation and adopted recent GPU implementations of computer vision algorithms and real-time techniques for robust estimation to speed up time-expensive subroutines in order to approach application on site.

Acknowledgments

This work has been funded by the "Zukunftsprogramm Schleswig-Holstein (2007-2013)" with funds from the European Commission (EFRE) and Land Schleswig-Holstein, Germany, as part of the Initiative KoSSE, project 122-09-048, and was supported by IBAK Helmut Hunger GmbH & Co. KG.

Literature

[BFG09] Burger, B.; Fiedler, M.; Gellrich, J.; Reuter, H.-P.: Schachtinspektion in neuer Qualität. Bauwirtschaftliche Information UmweltBau Nr. 1, 2009.

[CPT98] Cooper, D.; Pridmore, T. P.; Taylor, N.: Towards the Recovery of Extrinsic Camera Parameters from Video Records of Sewer Surveys. In: Machine Vision and Applications 11, p.53-63, 1998.

[EKR09] Esquivel, S.; Koch, R.; Rehse, H.: Reconstruction of Sewer Shaft Profiles from Fisheye-Lens Camera Images. In: Lecture Notes in Computer Science 5748, p.332-341, 2009.

[EKR10] Esquivel, S.; Koch, R.; Rehse, H.: Time Budget Evaluation for Image-Based Reconstruction of Sewer Shafts. To be published in: Proc. SPIE Photonics Europe '10, Brussels, Belgium, 2010.

[FB81] Fischler, M. A.; Bolles, R. C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: Communications of the ACM 24 (6), p.381-395, 1981.

[JU97] Julier, S. J.; Uhlmann, J. K.: A New Extension of the Kalman Filter to Nonlinear Systems. In: International Symposium on Aerospace/Defense Sensing, Simulation and Controls, 1997.

[HZ03] Hartley, R.; Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd Edition. Cambridge University Press, 2003.

[KBH08] Kannala, J.; Brandt, S. S.; Heikkilä, J.: Measuring and Modelling Sewer Pipes from Video. In: Machine Vision and Applications 19 (2), p.73-83, 2008.

[Ni05] Nistér, D.: Preemptive RANSAC for Live Structure and Motion Estimation. In: Machine Vision and Applications 16 (5), p.321-329, 2005.

[SBK08] Schiller, I.; Beder, C.; Koch, R.: Calibration of a PMD-Camera Using a Planar Calibration Pattern together with a Multi-Camera Setup. In: Proc. ISPRS Congress XXI., Beijing, 2008.

[TK91] Tomasi, C.; Kanade, T.: Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.

[TMH99] Triggs, B.; McLauchlan P.; Hartley R.; Fitzgibbon A.: Bundle Adjustment – A Modern Synthesis. In: Proc. ICCV'99 International Workshop on Vision Algorithms, Springer-Verlag, p.298-372, 1999.

[ZGF08] Zach, C.; Gallup, D.; Frahm, J.-M.: Fast Gain-Adaptive KLT Tracking on the GPU. In: Proc. CVPR '08 Workshop on Visual Computer Vision on GPUs, 2008.