Machine Vision and Applications manuscript No.
(will be inserted by the editor)

Juho Kannala · Sami S. Brandt · Janne Heikkilä

Measuring and Modelling Sewer Pipes from Video

Received: date / Accepted: date

Abstract This article presents a system for the automatic measurement and modelling of sewer pipes. The system recovers the interior shape of a sewer pipe from a video sequence which is acquired by a fish-eye lens camera moving inside the pipe. The approach is based on tracking interest points across successive video frames and posing the general structure-from-motion problem. It is shown that the tracked points can be reliably reconstructed despite the forward motion of the camera. This is achieved by utilizing a fish-eye lens with a wide field of view. The standard techniques for robust estimation of the two- and three-view geometry are modified so that they can be used for calibrated fish-eye lens cameras with a field of view less than 180 degrees. The tubular arrangement of the reconstructed points allows pipe shape estimation by surface fitting. Hence, a method for modelling such surfaces with a locally cylindrical model is proposed. The system is demonstrated with a real sewer video and an error analysis for the recovered structure is presented.

Keywords Structure from motion · 3D-Reconstruction · Omnidirectional vision · Visual inspection · Modelling

J. Kannala[1], S. S. Brandt[2,1], J. Heikkilä[1]

[1] Machine Vision Group,
University of Oulu, Finland
E-mail: {jkannala,jth}@ee.oulu.fi

[2] Laboratory of Computational Engineering,
Helsinki University of Technology, Finland
E-mail: sami.brandt@tkk.fi

1 Introduction

The condition assessment of sewer pipes is usually carried out by visual inspection of sewer video sequences. However, the manual inspection has a number of drawbacks such as subjectivity, varying standards and high costs. Therefore several approaches for automation of sewer surveys have been suggested. For example, automatic detection of pipe joints and surface cracks from digital sewer images has been investigated [33,4,28]. An idea of recovering the three-dimensional shape of a surveyed pipe from survey videos was presented in [5], where a method for determining the pose of the camera relative to the central axis of the pipe was additionally proposed. Different kinds of sewer robots have also been developed and some of them contain additional sensors besides the video camera [13,20]. While multisensoric robots provide additional information, they also lead to a more complex and expensive construction.

In this paper, we describe a method for recovering the shape of a sewer pipe solely from a video sequence which is acquired by a fish-eye lens camera.[1] Our approach is to track interest points across successive images and address the structure-from-motion (SFM) problem in the case of fish-eye image sequences. Despite the fact that SFM for perspective cameras is extensively studied there are not yet many real *omnidirectional* SFM results published for long image sequences [12]. One of the first such systems is presented in [21] where a catadioptric camera is used.

Another contribution of this paper is to describe a practical method for modelling tubular 3D structures. In order to model the bending of pipes we use a model which is concatenated from short cylindrical pieces and thereafter smoothed along the pipe. The fitting of cylindrical surfaces has been discussed for example in [24] and [31] but the difference to these approaches is that our approach minimizes a geometric cost function, is robust to outliers, and does not require the cylinder to be circular.

It should be noticed that the techniques proposed in this paper are not limited to the sewer monitoring application but they can be also used to recover scene structure from other video se-

[1] Part of this work has been published in [17].

quences taken by a calibrated fish-eye lens camera. Likewise, the modelling of tubular surfaces may be needed in other applications as well. In particular, we believe that the work presented here might be useful in certain medical applications. For example, there have already been some attempts to use computer vision techniques for structure recovery from endoscopic images [27], [3].

The structure of the paper is as follows. In Section 2, we give an overview of our system and, in Section 3, we describe the differences to conventional structure-from-motion approaches in more detail. The methods for error analysis are described in Section 4. The modelling of pipes is discussed in Section 5 and results of the experiments with a real sewer video are reported in Section 6.

2 Overview of the method

A typical sewer inspection system consists of a video camera and a remote controlled tractor. Such a system is illustrated in Fig. 1. The sewer robot we used had a fish-eye lens camera whose wide field of view makes it possible to obtain a high resolution scan of the whole pipe by a single pass. The wide field of view is essential in our application since it allows reliable structure recovery even when the camera is moving forward. In fact, it is well known that for perspective cameras forward motion can give poor reconstructions since corresponding rays between successive views are almost parallel for most of the field of view [15].

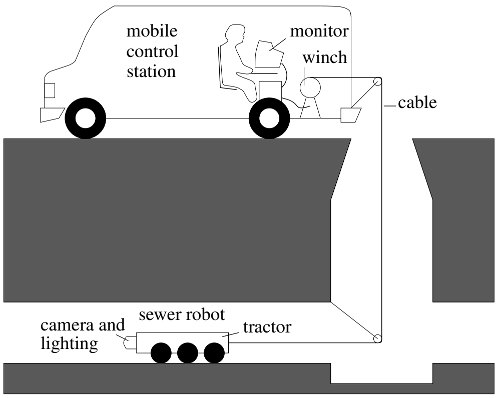


Fig. 1: A typical sewer inspection system

Our approach to structure recovery mainly follows the framework presented in [9]. However, since the usual pinhole camera model is not a valid approximation to a fish-eye lens several modifications are proposed. In the following, we briefly describe the different steps in our method.

2.1 Camera calibration

Although the modern approach to structure recovery is often uncalibrated [9], we adopt the traditional photogrammetric principle of camera calibration prior to measurements. One reason for this is the peculiarity of the fish-eye lens and the other is the requirement for high accuracy. In our case the calibration is done by viewing a planar calibration object [18]. Since the corresponding camera model is very flexible it is suitable for fish-eye lenses with severe distortion. The calibration gives the transformation $\mathcal{T}$ that warps the original image to a perspective one, i.e., transforms the fish-eye image coordinates $\mathbf{m}$ to $\tilde{\mathbf{x}} = \mathcal{T}(\mathbf{m})$, which are the normalized image coordinates of a pinhole camera.

2.2 Feature extraction, matching and tracking

The tracked features are interest points detected by the Harris corner detector [14], which is widely used in these kinds of applications. The experiments with sewer video sequences showed that there are plenty of such features in eroded concrete pipes. The detected interest points are initially matched between each successive image pair through normalized cross-correlation of the intensity neighbourhoods. Here the correlation threshold used for matching was 0.75 for windows of size 7×7.

The putative point correspondences contain almost unavoidably some false matches. In the tracking step, the aim is to use the geometric constraints between successive view pairs and view triplets to guide the matching and to discard the false matches. This requires some modifications to the usual way of estimating the multiple view geometry [9]. The modifications are needed to ensure a justified distribution of estimation error when the image correspondences are measured from the original fish-eye images and the two and three view relations, defined by the essential matrix and the trifocal tensor, hold between the warped images. A description of the proposed modifications is given in Section 3.

2.3 Reconstruction

The next step is to recover the structure by computing the three-dimensional coordinates of the tracked points. If enough interest points can be tracked and reconstructed, the arrangement of the corresponding three-dimensional points should be tubular allowing to estimate the shape of the pipe.

Here we use a hierarchical method that is similar to [10] and distributes the reconstruction error over the whole sequence. The idea is to compute the camera motion and the 3D points for each image triplet and then hierarchically registrate the triplets into longer sub-sequences which are bundle adjusted at each level. In contrast to [10], our approach is calibrated. This makes the algorithm simpler since then the registration of 3D points from two overlapping sub-sequences requires finding a 3D similarity transformation instead of a general projective transformation and hence the transform may be computed by a non-iterative algorithm [30].

2.4 Modelling

The final step is to make a 3D model of the pipe by fitting a surface to the reconstructed points. Here we use a piecewise cylindrical model since a straight cylinder would be too rigid model to capture the bending of pipes. In the fitting process the points are first divided into several sections along the pipe and a cylinder with elliptical cross-section is fitted to each section. Finally the parameters of cylindrical pieces are interpolated along the pipe so that the surface is smooth also in the main axis direction. The details of our approach are described in Section 5.

3 Geometry of fish-eye views and tracking

Let $\mathbf{m}_j^i$ be the measured coordinates of point correspondence i in view j. Given these measured correspondences and assuming that the image measurement errors obey a zero-mean isotropic Gaussian distribution, the Maximum Likelihood estimate for the camera motion is obtained by minimizing

$$\sum_i \sum_j \delta_{ij} d(\mathbf{m}_j^i, \hat{\mathbf{m}}_j^i)^2, \tag{1}$$

where $d(\cdot,\cdot)$ is the distance between two image points, δ_{ij} is either 1 or 0 indicating whether the point i is observed in view j, and $\hat{\mathbf{m}}_j^i$ are the estimated correspondences

$$\hat{\mathbf{m}}_j^i = \mathcal{P}(\hat{\mathbf{X}}^i, \boldsymbol{\theta}_j). \tag{2}$$

Here $\mathcal{P}$ is the imaging function of the fish-eye camera, $\boldsymbol{\theta}_j$ are the motion parameters in view j and $\hat{\mathbf{X}}^i$ represent the estimated 3D coordinates of point i. Since the camera is calibrated, the values of the internal camera parameters are known and the cost (1) should be minimized over the external camera parameters $\boldsymbol{\theta}_j$ and the 3D coordinates $\hat{\mathbf{X}}^i$. However, the direct minimization of (1) requires a good initialisation and does not tolerate false matches. Hence, we implemented the RANSAC algorithm for the robust estimation of camera motion between view pairs and triplets. The implementation follows the general recommendations in [15] but the adaptation to the fish-eye case is our own and is described in the following sections. We would like to point out that another possible implementation for robust motion estimation in the calibrated case is presented in [26].

3.1 Two views

Consider the case of two views, $j = \{1, 2\}$, and assume that there is a set of putative point correspondences, $\mathbf{m}_1^i \leftrightarrow \mathbf{m}_2^i$. The transformed coordinates are $\tilde{\mathbf{x}}_j^i = \mathcal{T}(\mathbf{m}_j^i)$ and the two view constraint between the transformed images is expressed by the essential matrix [15].

In RANSAC, we randomly select samples of seven point correspondences and each sample gives one or three candidates for the essential matrix [15]. Then, given an essential matrix candidate $\mathbf{E}$ and the transformed correspondences, $\tilde{\mathbf{x}}_1^i \leftrightarrow \tilde{\mathbf{x}}_2^i$, there is a non-iterative algorithm [15] for computing such points $\hat{\mathbf{x}}_1^i$ and $\hat{\mathbf{x}}_2^i$ that minimize the geometric distance

$$\sum_i d(\tilde{\mathbf{x}}_1^i, \hat{\mathbf{x}}_1^i)^2 + d(\tilde{\mathbf{x}}_2^i, \hat{\mathbf{x}}_2^i)^2 \tag{3}$$

in the transformed image plane subject to the constraint

$$\hat{\mathbf{x}}_2^{i\top} \mathbf{E}\, \hat{\mathbf{x}}_1^i = 0.$$

By transforming the points $\hat{\mathbf{x}}_j^i$ to the original image, one obtains the points

$$\hat{\mathbf{m}}_j^i = \mathcal{T}^{-1}(\hat{\mathbf{x}}_j^i), \tag{4}$$

which may be used as approximations to the optimal exact correspondences in (1). We use (4) to compute the distances

$$d(\mathbf{m}_1^i, \hat{\mathbf{m}}_1^i)^2 + d(\mathbf{m}_2^i, \hat{\mathbf{m}}_2^i)^2 \tag{5}$$

and classify the correspondences into inliers and outliers. In practice, it is important that this distance is measured in the original image instead of the transformed image plane since the transformation $\mathcal{T}$ is highly non-linear. Otherwise the distances near the border of field of view would get too large weight. As usual, the $\mathbf{E}$ which has most inliers is chosen and it gives our first estimate for the camera motion. For comparison, a 5-point method is used to solve $\mathbf{E}$ in [26] but it is also reported there that the stability of the 5-point method for forward motion is worse than that of the 7-point method.

Since the essential matrix may be parameterized by the rotation and translation parameters [15], the equation (4) implicitly defines the points $\hat{\mathbf{m}}_j^i$ as a function of the external camera parameters and the measured correspondences. Hence, by substituting the points (4) into (1) one obtains a cost function which is minimized over the 5 parameters of the essential matrix only. So, given an initial estimate for $\mathbf{E}$ we use the approximation (4) and refine our motion estimate by minimizing (1) using the inlier correspondences.

3.2 Uncertainty of the two-view geometry

The cost function in (1) has such a form that, as a by-product of the minimization, one can compute an estimate for the covariance of the parameters of the essential matrix [6]. The estimated covariance may be used to compute the epipolar envelopes which constrain the search region for new correspondences.

In detail, given a point $\mathbf{m}$ in the first image and the corresponding epipolar line $\mathbf{l}$ in the transformed image plane of the second image, one may use the covariance of the essential matrix in order to compute the first-order approximation for the covariance of the epipolar line [15]. Then, assuming that $\mathbf{l}$ is a random line obeying a Gaussian distribution with the mean at the estimated value and covariance $\boldsymbol{\Lambda}_{\mathbf{l}}$, the epipolar envelope is the conic

$$\mathbf{C} = \mathbf{l}\mathbf{l}^\top - k^2 \boldsymbol{\Lambda}_{\mathbf{l}}, \tag{6}$$

which represents an equal-likelihood contour bounding some fraction of all instances of $\mathbf{l}$ [15]. If $F_2(k^2)$ represents the cumulative χ_2^2 distribution and k^2 is chosen such that $F_2(k^2) = \alpha$, then a fraction α of all lines lie within the region bounded by $\mathbf{C}$.

We illustrate the estimated two view geometry in Fig. 2 where two successive images of a sewer video sequence are shown. We have chosen two points from the first image and plotted the corresponding epipolar curves and envelopes to the second image by transforming the epipolar lines and the hyperbolas (6) to the original fish-eye image. The yellow crosses in the second image are the narrowest points of the envelopes [2]. The narrow envelope of the vertical curve is the 95 % confidence interval used in our experiments to constrain the search region.

3.3 Three views

The two view constraint significantly reduces the occurrence of false matches but the three view constraint is even more effective. In the three-view case, we first robustly estimate the camera motion for view pairs (1,2) and (1,3). Then the only quantity that is left undetermined is the relative scale of the two translations. We use the RANSAC procedure to determine this ratio

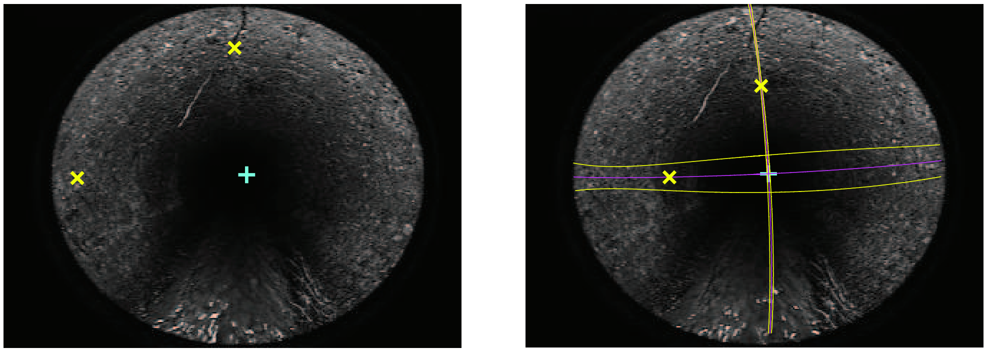


Fig. 2: Estimated epipolar geometry for two fish-eye images. Two points in the first image are chosen (yellow crosses) and their epipolar curves (magenta curves) are plotted to the second image. The yellow curves are the epipolar envelopes. The envelope of the horizontal curve is broad because a very large value of $k^2 = 1000$ was chosen in (6) in order to better illustrate the error bounds. The narrow confidence interval of the vertical curve is the 95 % envelope that corresponds to a value $k^2 = 5.99$.

from the three-view correspondences. At minimum only one additional sample correspondence needs to be drawn [16]. The distance measure used for the classification of inliers is

$$d(\mathbf{m}_1^i, \hat{\mathbf{m}}_1^i)^2 + d(\mathbf{m}_2^i, \hat{\mathbf{m}}_2^i)^2 + d(\mathbf{m}_3^i, \hat{\mathbf{m}}_3^i)^2, \tag{7}$$

where $\hat{\mathbf{m}}_1^i$ and $\hat{\mathbf{m}}_2^i$ are computed exactly as in (5) and $\hat{\mathbf{m}}_3^i$ is the point that is obtained by transferring the correspondence $\hat{\mathbf{m}}_1^i \leftrightarrow \hat{\mathbf{m}}_2^i$ to the third view with the trifocal point transfer.

The final estimate of the camera motion over each triple of views is refined by minimizing (1) over both the motion parameters and the 3D coordinates of the inliers. We additionally iterate between (i) least-squares fit to inliers and (ii) re-classification of inliers, until convergence. The estimated camera motion for each image triplet also provides a basis for the final reconstruction by bundle adjustment as described in Section 2.3.

The robustly estimated three view geometry is used to guide the matching when establishing the final correspondences. Since the geometric constraint discriminates the false matches, a weaker similarity threshold can be employed for the correlation windows. By accepting only such correspondences that are found from at least three successive images, the occurrence of false matches becomes unlikely. Utilizing the three-view constraint is essential in our application since the concrete surface of the pipe has fine repeating texture and the risk for false

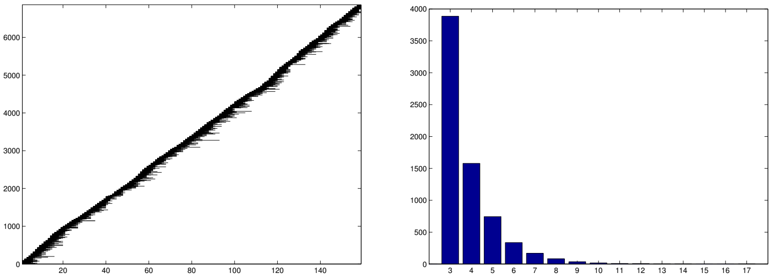


Fig. 3: Left: The tracked points through a sequence of 159 sewer images. Right: Histogram of the track lengths.

matches is high even after enforcing the two-view constraint. The final point tracks for a sequence of 159 sewer images are illustrated in Fig. 3. Since the camera constantly moves forward the tracks are short and most of them appear only in one triple of images.

4 Reconstruction and error analysis

4.1 Bundle adjustment

The final estimate for the structure and motion is obtained by global bundle adjustment minimizing (1). The cost function (1) can be written as

$$||\mathbf{y} - f(\mathbf{x})||^2, \tag{8}$$

where $\mathbf{y}$ is the measurement vector formed by stacking all the observed image points $\mathbf{m}_j^i$ into a single vector whose dimension N is two times the number of non-zero δ_{ij}:s in (1). The vector $\mathbf{x}$ contains both the structure and camera parameters and the function $f : \mathbb{R}^M \rightarrow \mathbb{R}^N$ computes the 2D points from 3D points and cameras. Often N and M are large which leads to a large-scale optimization problem. However, the structure of the problem is inherently sparse since individual point measurements are independent of each other. In our application additional sparseness arises from the fact that point tracks appear in only few views as illustrated in Fig. 3. Methods for solving sparse nonlinear least-squares problems are described in [15] and [29].

4.2 Error analysis

The uncertainty of estimated parameters is often evaluated by propagating the measurement covariance backwards to the parameter space using direct first-order analysis [15]. However, it is not always straightforward to use this technique for complex 3D reconstructions where the number of parameters is huge. The recent work [22] describes a method to deal with such high dimensional problems and this is the approach that is used also here.

The gauge constraint that we used in the uncertainty calculations is the symmetric camera-based gauge introduced in [22]. This gauge constraint handles all the cameras equally and is suitable for evaluating camera uncertainties in camera localization applications. In the gauge-free situation the function f in (8) is overparameterized so that each camera contributes six parameters and each point three parameters. Then the symmetric camera-based gauge constraint is defined by the following seven independent scalar equations [19]

$$\sum_j \mathbf{t}_j = \mathbf{0} \tag{9}$$

$$1 - \sum_j ||\mathbf{t}_j||^2 = 0 \tag{10}$$

$$\sum_j \mathbf{t}_j^* \times \mathbf{t}_j = \mathbf{0} \tag{11}$$

where $\mathbf{t}_j$ is the center of the j:th camera and $\mathbf{t}_j^*$ is the *estimate* of $\mathbf{t}_j$ provided by the bundle adjustment. Here (9) fixes the translation, (10) fixes the scale and (11) fixes the rotation. The 7 equations above can be written in the form $c(\mathbf{x}) = \mathbf{0}$ which defines a manifold of dimension $(M-7)$ in $\mathbb{R}^M$.

Let $\mathbf{x}^*$ be a point in the parameter space which minimizes (8) and satisfies $c(\mathbf{x}^*) = \mathbf{0}$. Further, let $\mathbf{J}_f$ and $\mathbf{J}_c$ be the Jacobians of f and c evaluated at $\mathbf{x}^*$. Assume that the measurement $\mathbf{y}$ obeys a zero-mean Gaussian distribution with the covariance $\sigma^2\mathbf{I}$ and mean $f(\mathbf{x}^*)$. Then the first order approximation for the parameter covariance around $\mathbf{x}^*$ under the constraint c is obtained by

$$\mathbf{C}_\mathbf{x} = \sigma^2 \mathbf{P}(\mathbf{J}_f^\top \mathbf{J}_f)^+ \mathbf{P}^\top, \tag{12}$$

where $\mathbf{P}$ is the linear projector on $\mathrm{Ker}(\mathbf{J}_c)$ parallel to $\mathrm{Ker}(\mathbf{J}_f)$ (Ker denotes the null space) [22]. Here the direct computation of the pseudoinverse $(\mathbf{J}_f^\top \mathbf{J}_f)^+$ is not possible because it is excessively large in size. However, the sparsity of $\mathbf{J}_f$ can be utilized in the computation as described in [22] where also the explicit form of $\mathbf{P}$ is given. An estimate for the variance σ^2 is obtained by $\sigma^2 = ||\mathbf{y} - f(\mathbf{x}^*)||^2/(N - (M-7))$.

The covariance matrix $\mathbf{C}_\mathbf{x}$ provides a basis for the uncertainty analysis of sewer reconstructions presented in Section 6. For example, the confidence ellipsoids of the camera centers and 3D points may be computed from diagonal 3×3 sub-blocks of $\mathbf{C}_\mathbf{x}$.

5 Modelling pipes

The reconstruction phase gives a set of 3D points which is used as a starting point for the modelling part. A characteristic feature of pipes is the cylindrical shape and hence we use cylinder as the basic building block of our model. Since the pipes may be bent, as shown in Fig. 6, we do not use a single straight cylinder but divide the 3D points into several short sections along the pipe and fit a cylinder to each section (see Fig. 4). The division can be easily done because the trajectory of the camera provides an estimate for the direction of the pipe. Here we take an interval of a few video frames and compute the planes that are orthogonal to the camera trajectory at the both ends of the interval. The points between the planes are then used as input for the cylinder fitting procedure. The cylinder fitting and the concatenation of cylinders into the final model are described in the following two sections.

5.1 Cylinder fitting

Cylinder fitting is an often encountered topic in 3D modelling and there are various different approaches to the problem [24], [31], [7], [1]. The closest work to that which is presented here is [7]. However, none of the previous approaches is directly applicable to our case since we have the following requirements for the fitting procedure: (a) robustness to outliers, (b) applicability to elliptical cylinders, and (c) minimization of geometrical cost function. Robustness to outliers is required since there may be erroneous points outside the pipe, which are caused by false matches that nevertheless satisfy the geometry constraints (see Fig. 6); or there may be something inside the pipe which does not belong to the model (see Fig. 4). We use a right elliptical cylinder model instead of a circular cylinder because the pipe may be compressed so that the cross-section is not a circle. Finally, minimizing geometric error instead of algebraic error is a statistically sound approach for surface fitting [15,7,1]. The approach in [7] also considers the above three requirements while the most important differences are in the chosen parameterization and in the robust initialization which are described in detail below.

Despite the good properties of geometric distance, many approaches for cylinder fitting use some approximation to the geometric distance or even algebraic distance [24], [31]. This may be partly due to the fact that, given a point, it is complicated to compute the corresponding closest point on an elliptical cylinder. However, the problem is reduced to computing the distance from a point to an ellipse [7] and there is a non-iterative algorithm for this which results in at most four solutions [11]. By utilizing this result we may compute the closest point non-iteratively and implement the geometric cylinder fitting without the time consuming nested iterations used in [1], where the fitting of general implicit surfaces was discussed. Nevertheless, in our application we do not minimize the squared geometric error due to its sensitivity to outliers but we use the Huber function of the geometric error [15].

In addition to the choice of a suitable cost function the choice of the parameterization must be done carefully. We have parameterized the right elliptical cylinder using eight parameters: three parameters for the position $\mathbf{r}$ (a point on the axis), three parameters for the orientation, $\mathbf{v}$, and two parameters for the lengths of the axes of the base ellipse, a and b. Hence, the cylinder is represented with 8-vector $\mathbf{c} = (\mathbf{r}^\top, \mathbf{v}^\top, a, b)^\top$. We fix the cylinder coordinate frame so that the z-axis is the axis of the cylinder and the x and y-axis are the axes of the base ellipse. Then the orientation of the cylinder is fixed by giving a rotation between the world frame and the cylinder frame.[2] An important difference to [7] is that our parameterization directly enforces the correct surface type and no additional constraints are needed.

Cylinder fitting by the nonlinear minimization of a robust cost function requires also a robust initialization which is not discussed in [7]. Given the data points and an initial estimate for the direction of the cylinder axis, we obtain the initialization for the base ellipse by the RANSAC-algorithm, i.e., we take 2D projection of the data perpendicular to the axis and fit an ellipse to the 2D points by using the method in [8] within the RANSAC-framework. In our case the initial axis direction is obtained from the camera trajectory.

The proposed approach for robust cylinder fitting is summarized in the following and an example of a fitted cylinder is shown in Fig. 4.

Procedure for robust cylinder fitting

The 3D data points $\mathbf{p}_i$, $i = 1, \ldots, n$, and an initial direction $\mathbf{d}$, $||\mathbf{d}|| = 1$, for the cylinder axis are given as input.

(i) Provide a routine A which, for a given 2D data point, computes the nearest point on a conic $\mathbf{C}$ [11]. By simply extending the routine A, provide a routine B which for a given 3D data point computes the nearest point on an elliptical cylinder $\mathbf{c}$.

(ii) Let $\mathbf{R}_\alpha$ be a rotation such that $\mathbf{R}_\alpha \mathbf{d} = (0\;0\;1)^\top$.

(iii) Compute the vectors $\mathbf{q}_i = \mathbf{R}_\alpha \mathbf{p}_i$ and denote their first two components by $\tilde{\mathbf{q}}_i$. Compute the median of the third component and denote it by z_0.

(iv) Robustly fit an ellipse to the points $\tilde{\mathbf{q}}_i$ using RANSAC: draw random samples of six points, fit by [8], and determine the outliers using the geometric distance.[3] Extract the five parameters of the ellipse: center point (x_0, y_0), rotation φ, and the lengths of the axes, (a_0,b_0). Compute the rotation parameters $\mathbf{v}_0$ which correspond to the composite rotation $\mathbf{R}_\varphi \mathbf{R}_\alpha$ where $\mathbf{R}_\varphi$ is a 3D rotation through an angle φ about the z-axis.

(v) Compute $\mathbf{r}_0 = \mathbf{R}_\alpha^\top (x_0, y_0, z_0)^\top$ and initialize the cylinder parameters by $\mathbf{c}_0 = (\mathbf{r}_0^\top, \mathbf{v}_0^\top, a_0, b_0)^\top$.

[2] We use the *angle-axis* representation [15, Appendix 4] for parameterizing the rotations, i.e., a 3-vector $\mathbf{v}$ represents a rotation through an angle $||\mathbf{v}||$ about the axis determined by $\mathbf{v}$.

[3] The distance threshold for outliers is estimated using the least median of squares method [32].

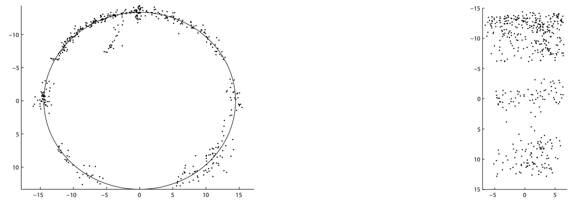


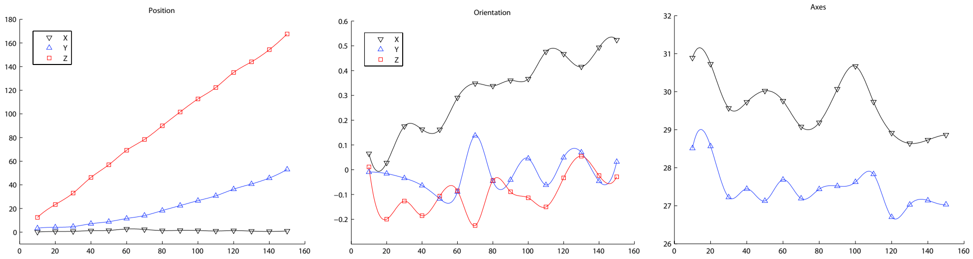
Fig. 4: Front and side views of the reconstructed points in a short section. The ellipse depicts the cross-section of the fitted cylinder. The points inside the pipe near the roof correspond to a root that is hanging from the roof.

(vi) Minimize $||\boldsymbol{\epsilon}||^2$ where the vector $\boldsymbol{\epsilon} = (\gamma_1 \boldsymbol{\epsilon}_1^\top, \ldots, \gamma_n \boldsymbol{\epsilon}_n^\top)^\top$ is formed by concatenating the residuals $\boldsymbol{\epsilon}_i = (\mathbf{p}_i - \hat{\mathbf{p}}_i)$ with weights $\gamma_i = C(||\boldsymbol{\epsilon}_i||)^{1/2}/||\boldsymbol{\epsilon}_i||$. Here $\hat{\mathbf{p}}_i$ are the closest points on the cylinder given by routine B and C is the Huber cost function [15]. The minimization is carried over the cylinder parameters using, for example, the Levenberg-Marquardt algorithm.

5.2 Tubular Model

After cylinder fitting the next step is to concatenate the cylinders in order to obtain a tubular model. However, the position of each cylindrical section has to be first fixed in the world frame. Since the cylinder model, used in Section 5.1, has an infinite length its position in the main axis direction is undefined, i.e., the parameters $\mathbf{r}$ may define any point on the axis. Hence, we update $\mathbf{r}$ to be such a point on the axis that when a plane perpendicular to the axis is positioned on $\mathbf{r}$ there are equal number of data points on both sides of the plane. This places the cylinder to the median of the data in the main axis direction. Then, the pipe is represented as a sequence of 8-vectors, $\mathbf{c}_1, \ldots, \mathbf{c}_m$, each vector corresponding to a one cylindrical section.

Our approach for modelling the pipe is to simply consider it as a set of eight one-dimensional sequences. Each of these sequences corresponds to a one parameter of the cylinder model as described above. Thus, the components of the vectors $\mathbf{c}_i$ can be seen as samples from the underlying one-dimensional sequences. In order to obtain a parametric representation for the pipe, we have to fit a model to the samples. There are several possibilities for the approximation of one-dimensional functions. Here we have chosen to use the cubic spline interpolation for all the eight parameters since it produces smooth approximations.



(a) (b) (c)

Fig. 5: The cylinder parameters $\mathbf{c} = (\mathbf{r}^\top, \mathbf{v}^\top, a, b)^\top$ are interpolated along the pipe; (a) the components of vector $\mathbf{r}$, (b) the components of vector $\mathbf{v}$, (c) the axis lengths a and b.

The spline interpolants for the parameters are illustrated in Fig. 5 where the symbols along the curves denote the original samples corresponding to the fitted cylinders. The modelled data is that shown in Fig. 6. In Fig. 5 there are eight parameters: 3 for the position (Fig. 5(a)), 3 for the orientation (Fig. 5(b)) and 2 for the axes of the base ellipse (Fig. 5(c)). The orientation parameters describe the rotation between the world frame and the cylinder frame. The topmost curve in Fig. 5(b) corresponds to the x-component of the parameter $\mathbf{v}$ and its increase reflects the bending of the pipe illustrated in Fig. 6(b). Further, Fig. 5(c) shows that the two axes of the cross-sectional ellipse have clearly different lengths which indicates that an ellipse is indeed a better model than a circle. Here the major axis of the ellipse is almost horizontal along the whole pipe. Finally, the obtained tubular model is illustrated in Fig. 7 where the sample density, corresponding to the plotted cylindrical sections, is twice the one that was used in fitting.

6 Results

We experimented one sewer video sequence scanned in an eroded concrete pipe with a Watec 221S color camera and a fish-eye lens whose nominal field of view was $180°$. The uncompressed digital video was captured from an analog NTSC video signal at a resolution of 320×240. The image sequence used in the experiments contained 159 fish-eye views and covered about two meters of the pipe. The sequence was made by including every fifth frame from a video clip which was taken while the camera was moving approximately at a constant velocity. The color images were transformed to 8-bit grayscale images before further processing.

The processing of the test sequence took about nine hours using our current Matlab implementation. The time calculations for the different steps were as follows: initial matching $3500s$, motion estimation for view pairs and triplets $22000s$, bundle adjustment $4700s$, and modelling $580s$.

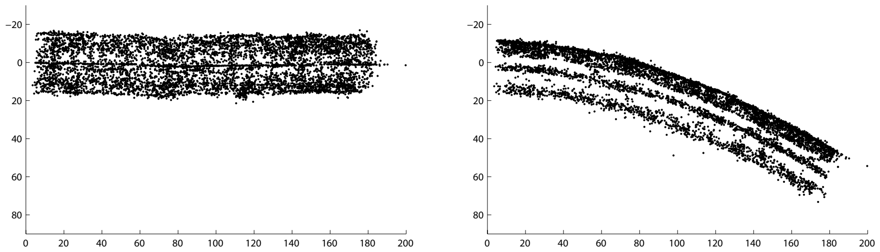
6.1 Reconstruction

The obtained reconstruction is illustrated in Fig. 6 and it was computed using the techniques described in Sections 3 and 4. The final estimates for the structure and motion were obtained via global bundle adjustment which involved 159 views and 6864 points. The sparse structure of the large-scale optimization problem was utilized in the nonlinear minimization [29, 15]. After the optimization the root-mean-squared (RMS) residual error was 0.18 pixels which is a reasonably small value. In addition, the maximum residual error was only 1.3 pixels which indicates that all the tracks agreed with the motion. Hence, the conventional least-squares cost function is justified here since the outliers were already removed in the tracking step.

Visual inspection of the recovered structure allows many interesting observations. First of all, the arrangement of the points is clearly tubular as shown in Figs. 4, 6 and 7. In addition, some small details, such as the root in Fig. 4 and the crack in Fig. 7, are clearly visible. This high visual quality is achieved despite the low resolution video, forward motion and bad lighting conditions inside the pipe. Further, there are few reconstructed points in the bottom part of the pipe since it is difficult to find correspondences from the water region. Fig. 4 also shows that the cross-section is more a flattened ellipse than a circle. This finding is consistent with the crack on the roof since such a crack typically indicates that the concrete pipe is vertically compressed and has a risk of collapse.

6.2 Uncertainty

A method for approximating the covariance of the estimated structure and camera parameters was described in Section 4.2. The covariance matrix $\mathbf{C}_\mathbf{x}$ in (12) provides a basis for the



(a)

(b)

Fig. 6: Top and side views of the reconstructed points for a sequence of 159 images. (a) xz-plane, (b) yz-plane

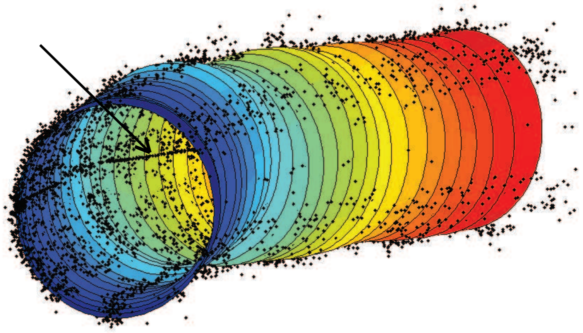


Fig. 7: The reconstructed points and the tubular model. The roof of the pipe is on the left. The sequence of interest points found from the crack on the roof is pointed by the arrow.

uncertainty analysis of the reconstruction. Let $\mathbf{C}_k$ be the 3×3 diagonal block of $\mathbf{C}_\mathbf{x}$ which corresponds to the parameters of the k:th camera center. Then the 90 $\%$ confidence ellipsoid for this camera center is given by $E_k(\mathbf{t}) = \{\mathbf{t} \in \mathbb{R}^3 | (\mathbf{t} - \mathbf{t}_k)^\top \mathbf{C}_k^{-1} (\mathbf{t} - \mathbf{t}_k) \leq F_3^{-1}(0.90)\}$, where F_3 is the cumulative χ^2 distribution with 3 degrees of freedom. The uncertainty ellipsoids for the reconstructed 3D points are obtained in a similar way.

The 90 $\%$ confidence ellipsoids for camera centers are illustrated in Fig. 8. One may observe that the uncertainty of cameras is larger at the ends of the sequence than in the middle. This seems natural since all the point tracks break at the ends. The reliability of the illustrated uncertainty bounds was additionally verified with simulations. The reconstruction in Fig. 6 was used as a ground truth, synthetic noise was added to the exact image measurements and bundle adjustment was carried out. The performed 20 trials gave results that were consistent with the analytical error bounds. For example, the estimated camera centers for all the trials were located inside the ellipsoids shown in Fig. 8.

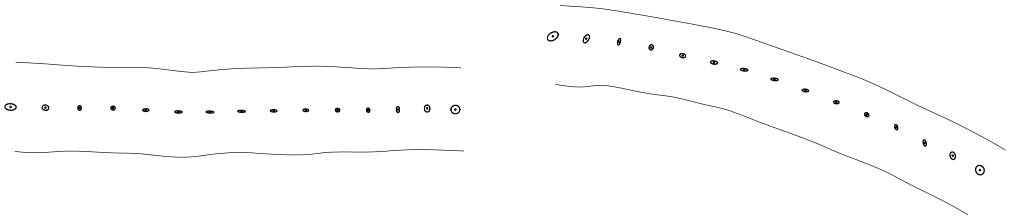


Fig. 8: Top and side views of the 90 % confidence ellipsoids for 15 cameras. The uncertainty analysis was performed using all cameras but only the cameras $k = 10, 20, \ldots, 150$ are shown for clarity.

Table 1: Quartiles of the semimajor axis of the confidence ellipsoids for points and cameras. The diameter of the pipe is 100.

	a_0	$a_{1/4}$	$a_{1/2}$	$a_{3/4}$	a_1
cameras	3.0	3.7	4.4	5.1	10.1
points	2.1	4.4	5.0	6.3	320

The quartiles of the semimajor axis of the 90 % confidence ellipsoids are shown in Table 1 for both points and cameras. Here a_0 is the smallest value, $a_{1/4}$ is the first quartile, $a_{1/2}$ is the median, $a_{3/4}$ is the third quartile, and a_1 is the largest value. The tabulated values are scaled so that the estimated average diameter of the pipe has a value 100, a value which is about 20 times larger than the median values in Table 1. It can also be seen that the localization uncertainty for most of the points is not much higher than that for the cameras. The maximum length of the confidence ellipsoid semiaxis is large for points (320) but there are only 14 points for which the ellipsoid semiaxis length exceeds the value 20. Note that the uncertainties of points are observed in the camera-based gauge which usually magnifies the uncertainty of the structure parameters [29].

In summary, the uncertainty of reconstructed points is small despite the short point tracks and the small baseline between successive cameras. The stability of the reconstruction is probably due to the fish-eye lens and its wide field of view. Similar advantages of omnidirectional cameras have been observed also by others [25]. In addition, it is interesting to qualitatively compare the sewer reconstruction to the corridor reconstruction presented in [23]. The corridor sequence in [23] has also a forward motion but the images are captured with a conventional camera. There the field of view covers only the frontal sector and many points have a high uncertainty in the reconstruction unlike in our case.

6.3 Modelling

In the modelling part the reconstructed points in Fig. 6 were divided into 15 sections and an elliptical cylinder was fitted to each section as described in Section 5. The fitting is automatic and robust, robustness is needed since all the points are not located on the surface of the pipe. For example, there are some erroneous points outside the pipe which agree with the motion constraints but do not correspond to any real observed point. The median distance of points from the surface of the corresponding cylinder was 1.2 whereas the average diameter of the pipe was 100. On average, the distances of points from the surface were smaller than the semimajor axis of their 90 % confidence ellipsoids.

After acquiring the cylindrical pieces their parameters were interpolated along the pipe as illustrated in Fig. 5. The obtained tubular model is illustrated in Fig. 7. There is a trade-off between the length of cylindrical sections and the flexibility of the model. Using short sections allows a bendy pipe but the cylinder fitting is more stable for longer sections. A suitable density of sections depends on the number of points and their distribution on the surface as well as on the type of the pipe. Here the proper number of sections was chosen manually.

7 Conclusion

This article describes a system for the measurement and modelling of sewer pipes from fish-eye image sequences. The experiments showed that the structure of a sewer pipe can be recovered solely from a video sequence that is scanned by a single pass through the pipe. The presented error analysis suggests that a relatively low uncertainty of reconstruction can be achieved despite forward motion of the camera. In addition, a method for modelling tubular 3D structures with a locally cylindrical model was proposed, and the robustness of the modelling approach was demonstrated in experiments with real data. The presented techniques are not necessarily limited to the sewer monitoring application but could be likewise utilized in other applications. In particular, since our camera model [18] can flexibly model various types of lenses and cameras with wide field of view provide advantages over the conventional ones, this kind of omnidirectional structure-from-motion techniques are expected to be useful in many robotic vision applications.

Acknowledgements We are grateful to Technical Research Centre of Finland and DigiSewer Productions Ltd. for cooperation and providing the sewer video sequences. The financial support by the Finnish Funding Agency for Technology and Innovation (TEKES) and the Graduate School in Electronics, Telecommunication and Automation (GETA) is gratefully acknowledged.

References

1. Ahn, S.J., Rauh, W., Cho, H.S., Warnecke, H.J.: Orthogonal distance fitting of implicit curves and surfaces. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 620–638 (2002)
2. Brandt, S.S.: On the probabilistic epipolar geometry. In: Proc. BMVC, pp. 107–116 (2004)
3. Burschka, D., Li, M., Taylor, R., Hager, G.D.: Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery. In: Proc. MICCAI, pp. 413–421 (2004)
4. Chae, M.J., Abraham, D.M.: Neuro-fuzzy approaches for sanitary sewer pipeline condition assessment. J. Comput. Civil Eng. **15**(1), 4–14 (2001)
5. Cooper, D., Pridmore, T.P., Taylor, N.: Towards the recovery of extrinsic camera parameters from video records of sewer surveys. Mach. Vis. and Appl. **11**, 53–63 (1998)
6. Csurka, G., Zeller, C., Zhang, Z., Faugeras, O.: Characterizing the uncertainty of the fundamental matrix. Comput. Vis. Image Underst. **68**(1), 18–36 (1997)
7. Faber, P., Fisher, B.: A buyer's guide to euclidean elliptical cylindrical and conical surface fitting. In: Proc. BMVC, pp. 521–530 (2001)
8. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least square fitting of ellipses. IEEE Trans. Pattern Anal. Mach. Intell. **21**(5), 476–480 (1999)
9. Fitzgibbon, A.W., Zisserman, A.: Automatic 3D model acquisition and generation of new images from video sequences. In: Proc. European Signal Processing Conference, pp. 1261–1269 (1998)
10. Fitzgibbon, A.W., Zisserman, A.: Automatic camera recovery for closed or open image sequences. In: Proc. ECCV, pp. 311–326 (1998)
11. Forsyth, D., Ponce, J.: Computer Vision, A Modern Approach. Prentice Hall (2003)
12. Geyer, C., Daniilidis, K.: Structure and motion from uncalibrated catadioptric views. In: Proc. CVPR, pp. 279–286 (2001)
13. Gooch, R.M., Clarke, T.A., Ellis, T.J.: A semi-autonomous sewer surveillance and inspection vehicle. In: Proc. IEEE Intelligent Vehicles, pp. 64–69 (1996)
14. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference (1988)

15. Hartley, R., Zisserman, A.: Multiple View Geometry, 2 edn. Cambridge (2003)

16. Kannala, J.: Measuring the shape of sewer pipes from video. Master's thesis, TKK (2004)

17. Kannala, J., Brandt, S.S.: Measuring the shape of sewer pipes from video. In: Proc. MVA (2005)

18. Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Trans. Pattern Anal. Mach. Intell. **28**(8), 1335–1340 (2006)

19. Kraus, K.: Photogrammetry, volume 2: Advanced methods and applications, Ferd. Dummler's Verlag, Bonn (1997)

20. Kuntze, H.B., Haffner, H.: Experiences with the development of a robot for smart multisensoric pipe inspection. In: Proc. IEEE Robotics and Automation, pp. 1773–1778 (1998)

21. Lhuillier, M.: Automatic structure and motion using a catadioptric camera. In: Proc. OMNIVIS (2005)

22. Lhuillier, M., Perriollat, M.: Uncertainty ellipsoids calculations for complex 3D reconstructions. In: Proc. IEEE Robotics and Automation, pp. 3062–3069 (2006)

23. Lhuillier, M., Quan, L.: A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 418–433 (2005)

24. Lukács, G., Martin, R., Marshall, D.: Faithful least-squares fitting of spheres, cylinders, cones and tori for reliable segmentation. In: Proc. ECCV, pp. 671–686 (1998)

25. Mičušík, B., Pajdla, T.: Structure from motion with wide circular field of view cameras. IEEE Trans. Pattern Anal. Mach. Intell. **28**(7), 1135–1149 (2006)

26. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 756–770 (2004)

27. Schmidt, J., Vogt, F., Niemann, H.: Nonlinear refinement of camera parameters using an endoscopic surgery robot. In: Proc. MVA, pp. 40–43 (2002)

28. Sinha, S.K., Fieguth, P.W.: Morphological segmentation and classification of underground pipe images. Machine Vision and Applications **17**, 21–31 (2006)

29. Triggs, B., McLauchlan, P.F., Hartley, R., Fitzgibbon, A.: Bundle adjustment - a modern synthesis. Lecture Notes in Computer Science **1883**, 298–372 (2000)

30. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. **13**(4), 376–380 (1991)

31. Werghi, N., Fisher, R., Robertson, C., Ashbrook, A.: Modelling objects having quadric surfaces incorporating geometric constraints. In: Proc. ECCV, pp. 185–201 (1998)

32. Xu, G., Zhang, Z.: Epipolar Geometry in Stereo, Motion and Object Recognition. Kluwer (1996)

33. Xu, K., Luxmoore, A.R., Davies, T.: Sewer pipe deformation assessment by image analysis of video surveys. Pattern Recognit. **31**(2), 169–180 (1998)