Visual Pipe Mapping with a Fisheye Camera

Peter Hansen, Hatem Alismail, Peter Rander and Brett Browning

CMU-CS-QTR-116
CMU-TR-RI-13-02

February 1, 2013

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

This publication was made possible by NPRP grant #08-589-2-245 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

Keywords:
Robotics, computer vision, pipe inspection, LNG, 3D mapping, visual mapping, visual odometry, SLAM, sparse bundle adjustment, structure from motion, fisheye.

Abstract

We present a vision-based mapping and localization system for operations in pipes such as those found in Liquified Natural Gas (LNG) production. A forward facing, fisheye camera mounted on a prototype robot collects imagery as it is tele-operated through a pipe network. The images are processed offline to estimate camera pose and sparse scene structure where the results can be used to generate 3D renderings of the pipe surface. The method extends state of the art visual odometry and mapping for fisheye systems to incorporate geometric constraints based on prior knowledge of the pipe components into a Sparse Bundle Adjustment framework. These constraints significantly reduce inaccuracies resulting from the limited spatial resolution of the fisheye imagery, limited image texture, and visual aliasing. Preliminary results are presented for a dataset collected in fiberglass pipe network which demonstrate the validity of the approach.

Contents

1 INTRODUCTION

Pipe inspection is a critical task to a number of industries, including Natural Gas production where pipe surface structure changes at the scale of millimeters are of concern. In this work, we report on the development of a fisheye visual odometry and mapping system for an in-pipe inspection robot (e.g. [11]) to produce detailed, millimeter resolution 3D surface structure and appearance maps. By registering maps over time, changes in structure and appearance can be identified, which are both cues for corrosion detection. Moreover, these maps can be imported into rendering engines for effective visualization or measurement and analysis.

In prior work [4], we developed a verged perspective stereo system capable of measuring accurate camera pose and producing dense sub-millimeter resolution maps. A limitation of the system was the inability of the camera to image the entire inner surface of the pipe. Here, we address this issue by using a forward-facing wide-angle fisheye camera mounted on a small robot platform, as shown in figure 1. This configuration enables the entire inner circumference to be imaged from which full coverage appearance maps can be produced. Figure 1 also shows the constructed pipe network used in the experiments, and sample images from the fisheye camera. The extreme lighting variations evident in the sample images pose significant challenges during image processing, as discussed in section 3.

Our system builds from established visual odometry and multiple view techniques for central projection cameras [8, 6, 12, 10, 9]. Binary thresholding, morphology and shape statistics are first used to classify straight sections and T-intersection. Pose and structure results are obtained for each straight section using a sliding window Sparse Bundle Adjustment (SBA) and localized straight cylinder fitting/regularization within the window. Fitting a new straight cylinder each window allows some degree of gradual pipe curvature to be modeled, e.g. sag in the pipes. While more generalized pipe models may be more suited for this purpose, for example cubic spline modeling of the pipe axis, computing the cylinder fitting regularization terms (distance of scene points to cylinder) could be prohibitively expensive. After processing each straight section, results for the T-intersections are obtained using SBA and a 2-cylinder intersection fitting/regularization – the 2 cylinders are the appropriate straight sections of the pipe network. As a final step, the pose and structure estimates are used to produce a dense point cloud rendering of the interior surface of the pipe network.

Results are presented in section 4 which show the visual odometry, sparse scene reconstruction, and 3D point cloud renderings for a one-loop dataset collected in our pipe network. These preliminary results illustrate the validity of the proposed system.

2 FISHEYE CAMERA

We use a $360^\circ \times 190^\circ$ angle of view Fujinon fisheye lens fitted to a $1280pix \times 960pix$ resolution CCD firewire camera. Image formation is modeled using a central projection polynomial mapping. A scene point $\mathbf{X}_i$ projects to a coordinate $\boldsymbol{\eta}(\theta, \phi) = \mathbf{X}_i / ||\mathbf{X}_i||$ on the camera's unit view sphere centered at the single effective viewpoint $(0, 0, 0)^T$. The angles θ, ϕ are, respectively, colatitude and longitude. The projected fisheye image coordinates are

$$\mathbf{u}(u, v) = \begin{bmatrix} (k_1\theta + k_2\theta^3 + k_3\theta^4 + k_4\theta^5)\cos\phi + u_0 \\ (k_1\theta + k_2\theta^3 + k_3\theta^4 + k_4\theta^5)\sin\phi + v_0 \end{bmatrix}, \quad (1)$$

where $\mathbf{u_0}(u_0, v_0)$ is the principal point. Multiple images of a checkerboard calibration target with known Euclidean geometry were collected, and the model parameters fitted using a non-linear minimization of the sum of squared checkerboard grid point image reprojection errors.

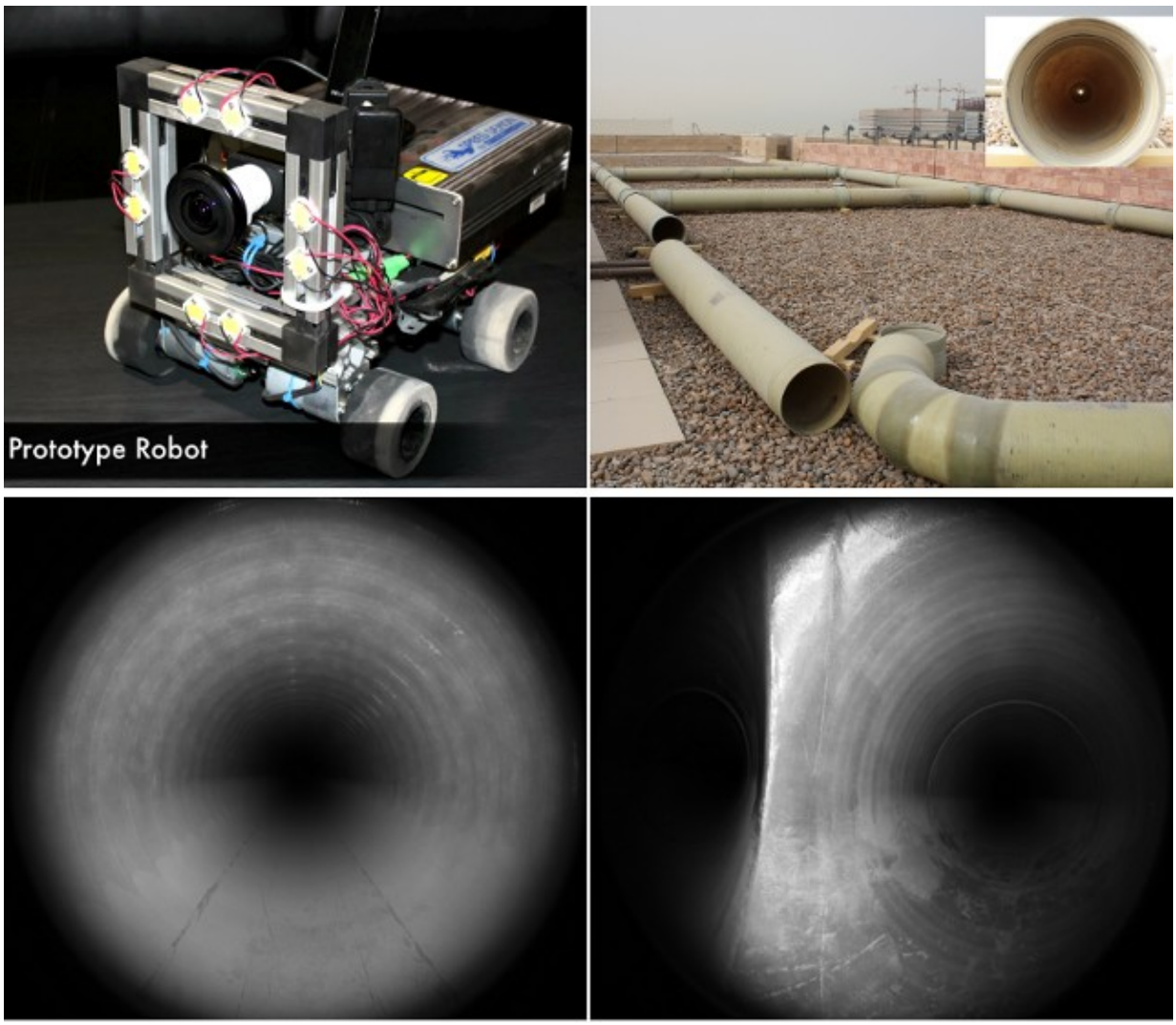


Figure 1: The prototype robot retrofitted with a forward facing fisheye camera. Images are logged to the onboard computer as the robot traverses the $400mm$ (16 inch) internal diameter fiberglass pipe network. Samples images in a straight section and T-intersection are shown. All lighting was provided from 8 high intensity LEDs surrounding the camera.

3 VISUAL ODOMETRY AND MAPPING

The visual odometry (VO) and mapping procedure is briefly summarized as follows:

A. Perform feature matching/tracking with keyframing.

B. Divide images into straight sections and T-intersections.

C. Obtain VO/structure estimates for each straight section using a sliding window SBA and localized straight cylinder fitting/regularization.

D. Obtain VO/structure estimates for each T-intersection using a 2 cylinder T-intersection model. This step effectively merges the appropriate straight sections.

The visual odometry steps (C and D) use different cylinder fitting constraints to obtain scene structure errors included as a regularization error in SBA. As previously mentioned, we have observed this to be a critically important step which significantly improves the robustness and accuracy of the visual odometry and scene reconstruction estimates in the presence of[1]: limited spatial resolution from the fisheye camera; feature location noise due to limited image texture and extreme lighting variations; and an often high percentage of feature tracking outliers due again to limited image texture and visual aliasing. At present an average *a priori* metric measurement of pipe radius r is used during cylinder fitting. Cylinder fitting with a supplied pipe radius also resolves monocular visual odometry scale ambiguity.

3.1 Feature Tracking

An efficient region-based Harris detector [5] based on the implementation in [10] is used to find a uniform feature distribution in each image. The image is divided into 2×2 regions, and the strongest $N = 200$ features per region are retained. Initial temporal correspondences are found using cosine similarity matching of 11×11 grayscale template descriptors for each feature. Each of these 11×11 template descriptors is interpolated from a 31×31 region surrounding the feature. Five-point relative pose [8] and RANSAC [3] are used to remove outliers and provide an initial estimate of the essential matrix E. We experimented with multiple scale-invariant feature detectors/descriptors (e.g. SIFT [7], SURF [1]), but observed no significant improvements in matching performance.

For all unmatched features in the first image, a guided Zero-mean Normalized Cross Correlation (ZNCC) is applied to find their pixel coordinate in the second image. Here, guided refers to a search within an epipolar *region* in the second image. Since we implement ZNCC in the original fisheye imagery, we back project each integer pixel coordinate to a spherical coordinate $\boldsymbol{\eta}$, and constrain the epipolar search regions using $\boldsymbol{\eta_2}^T E \, \boldsymbol{\eta_1} < thresh$ — the subscripts denote image 1 and 2. As a final step we implement image keyframing, selecting only images separated by a minimum median sparse optical flow magnitude or minimum percentage correspondences. Both minimums are selected empirically.

Figure 2 shows examples of the sparse optical flow vectors between keyframes in both a straight section and T-intersection. Features are 'tracked' across multiple frames by recursively matching using the method described.

3.2 Image classification: straight vs. T-intersection

To classify each image as belonging to a straight section or T-intersection, the image resolution is first reduced by sub-sampling pixels from every second row and column. A binary thresholding is applied to extract dark blobs within the cameras field of view, followed by binary erosion and clustering of the blobs. The largest blob is selected and the the second moments of area L and L_p computed about the the blob centroid and principal point, respectively. An image is classified as straight if the ratio L_p/L is less than an empirical threshold; we expect to see a large round blob near the center of images in straight sections. Figures 3a through 3c show the blobs extracted in three sample images and the initial classification of each image. After initial classification, a temporal filtering is used to remove false positive classification, as illustrated in figure 3d. This filtering enforces a minimum straight/T-intersection cluster size.

[1]Fiberglass pipes are significantly more challenging to process than steel as they contain reduced image texture and exhibit greater specular reflection. Steel pipes were unable to be used in the pipe network used for testing.

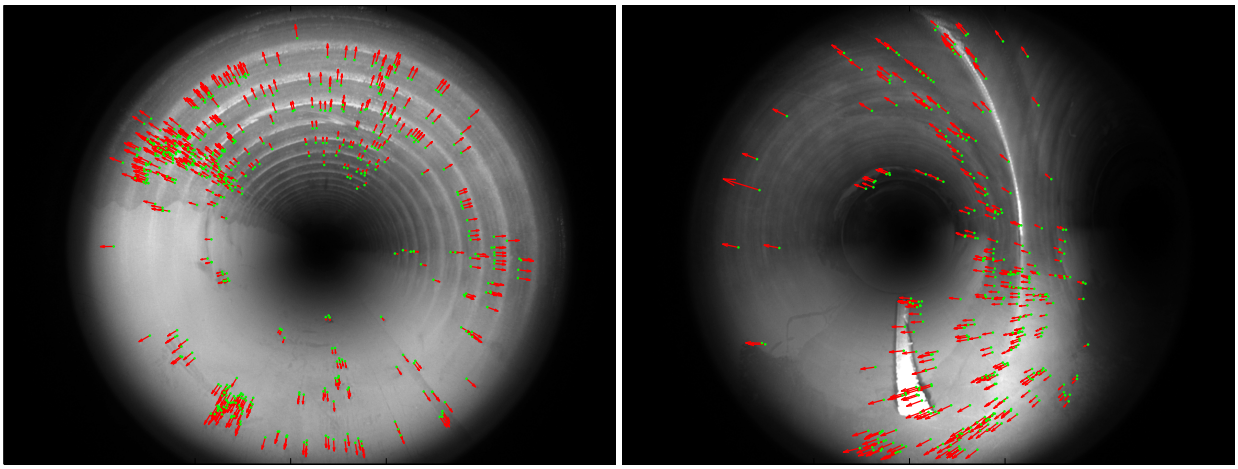
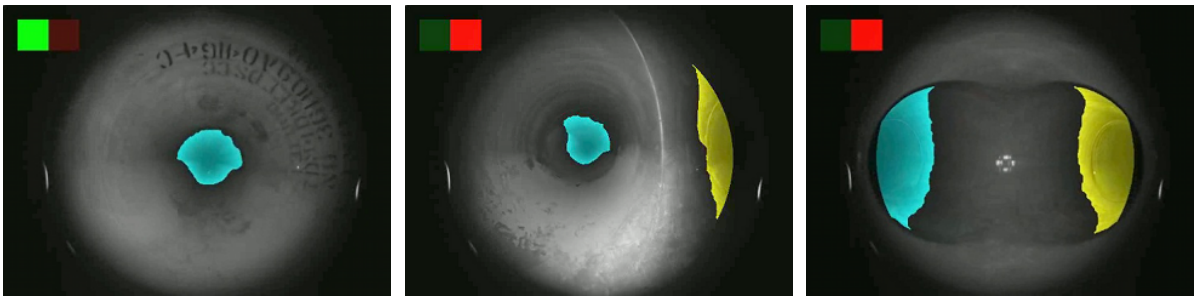
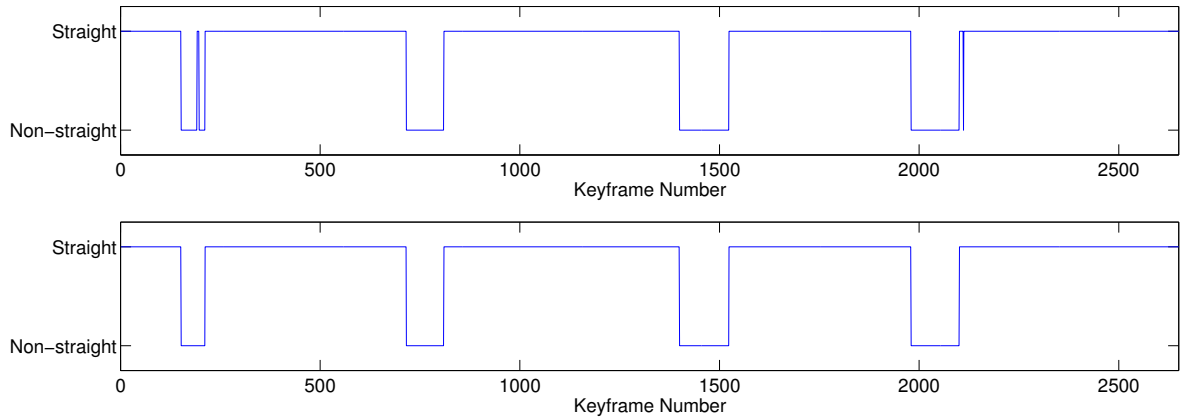


Figure 2: Sparse optical flow vectors in a straight section (left) and T-intersection (right) obtained using a combination of Harris feature matching and epipolar guided ZNCC.



(a) Straight section **(b) T-intersection** **(c) T-intersection**



(d) Initial classification (top), and after applying temporal filtering (bottom). Each of the four T-intersection clusters is a unique T-intersection in the pipe network.

Figure 3: Straight section and T-intersection image classification. Example initial classifications (a-c), and the classification of all keyframes before and after temporal filtering (d).

3.3 Straight VO: Sliding Window SBA / local straight cylinder

For each new keyframe, the feature correspondences are used to estimate the new camera position and scene points. This includes using Nister's 5-point algorithm to obtain an initial unit-magnitude pose change estimate, optimal triangulation to reconstruct the scene points [6], and prior scene coordinates to resolve relative scale.

After every 50 keyframes, a modified sliding window SBA is implemented which includes a localized straight cylinder fitting used to compute a scene point regularization error. A 100 keyframe window size is used which, on average, equates to a segment of pipe approximately one meter in length. This SBA is a multi-objective least squares minimization of image reprojection errors $\epsilon_{\mathbf{I}}$ and scene point errors $\epsilon_{\mathbf{X}}$. An optimal estimate of the camera poses P and scene points $\mathbf{X}$ in the window, as well as the fitted cylinder C are found which minimize the combined error ϵ:

$$\epsilon = \epsilon_{\mathbf{I}} + \epsilon_{\mathbf{X}}. \tag{2}$$

The image reprojection error $\epsilon_{\mathbf{I}}$ is the sum of squared differences between all valid feature observations $\mathbf{u}$ and reprojected scene point coordinates $\mathbf{u}'$:

$$\epsilon_{\mathbf{I}} = \sum_i ||\mathbf{u}_i - \mathbf{u}'_i||^2, \tag{3}$$

where $\mathbf{u}(u, v)$ and $\mathbf{u}'(u', v')$ are both inhomogeneous fisheye image coordinates.

The scene point error term $\epsilon_{\mathbf{X}}$ is a scalar weighted sum of squared errors between the optimized scene point coordinates $\mathbf{X}$ and a fitted straight cylinder. The origin of the cylinder is the first camera pose $P_m = [R_m|\mathbf{t}_m]$ in the sliding window, and is parameterized using 4 degrees of freedom:

$$\begin{aligned} C &= [\tilde{R}\,|\,\tilde{\mathbf{t}}] \\ &= [R_X(\gamma)\,R_Y(\beta)\,|\,(t_X, t_Y, 0)^T], \end{aligned} \tag{4}$$

where R_A denotes a rotation about the axis A, and t_A denotes a translation in the axis A. Each scene point coordinate $\mathbf{X}_i$ maps to a coordinate $\tilde{\mathbf{X}}_i$ in the cylinder coordinate frame using

$$\tilde{\mathbf{X}}_i = \tilde{R}\,(R_m\mathbf{X}_i + \mathbf{t}_m) + \tilde{\mathbf{t}}. \tag{5}$$

The regularization error $\epsilon_{\mathbf{X}}$ is

$$\epsilon_{\mathbf{X}} = \tau \sum_i \left(\sqrt{\tilde{X}_i^2 + \tilde{Y}_i^2} - r \right)^2, \tag{6}$$

where the pipe radius r is supplied. Adjusting the scalar τ controls the trade-off between the competing error terms $\epsilon_{\mathbf{I}}$ and $\epsilon_{\mathbf{X}}$. We use an empirically selected value $\tau = 100/r$.

As noted previously, there are frequently many feature correspondence outliers resulting from the challenging raw imagery. To minimize the influence of outliers, a Huber weighting is applied to individual error terms before computing $\epsilon_{\mathbf{I}}$ and $\epsilon_{\mathbf{X}}$. Outliers are also removed at multiple stages (iteration steps) using Median Absolute Deviation of the set of all signed Huber weighted errors $\mathbf{u} - \mathbf{u}'$.

3.4 T-intersections

The general procedure for processing the T-intersections is illustrated in figure 4. After processing each straight section, straight cylinders are fitted to the scene points in the first and last 1 meter segment (figure 4a). In both cases, these cylinders are fitted with respect to the first and last camera poses as the origins, respectively, using the parameterization in (4).

As illustrated in fig. 4b, a T-intersection is modeled as two intersecting straight cylinders; the red cylinder intersects the blue cylinder at a unique point $\mathcal{I}$. Let P_r be the first/last camera pose in a red section, and C_r be the cylinder fitted with respect to this camera as the origin. Similarly, let P_b be the last/first camera pose in a blue section, and C_b be the cylinder fitted with respect to this camera as the origin. The parameters ζ_r and ζ_b are rotations about the axis of the red and blue cylinders, and l_r and l_b are the signed distances of the cylinder origins $\mathcal{O}(C_r)$ and $\mathcal{O}(C_b)$ from the intersection point $\mathcal{I}$. Finally, ϕ is the angle of intersection between the two cylinder axes in the range $0^\circ < \phi < 180^\circ$. These parameters fully define the change in pose Q between P_b and P_r, and ensure that the 2 cylinder axes intersect at a single unique point $\mathcal{I}$. Letting

$$D = p\left([R_Z(\zeta_r)|(0,0,l_r)], [R_Z(\zeta_b)\, R_Y(\phi)|(0,0,l_b)]^T]\right), \tag{7}$$

where $p(b,a)$ is a projection a followed by b, and R_A is a rotation about axis A, then

$$Q = p\left(inv(C_r), p(D, C_b)\right), \tag{8}$$

where $inv(C_r)$ is the inverse projection of C_r.

SBA is used to optimize all camera poses P_T in the T-intersection between P_r and P_b, as well as all new scene points $\mathbf{X}$ in the T-intersection, and the T-intersection model parameters $\Phi(\zeta_r, l_r, \zeta_b, l_b, \phi)$. Again, the objective function minimized is the same form as (2), which includes an image reprojection error (3) and scene fitting error (6). The same value $\tau = 100/r$, robust cost function, and outlier removal scheme are also used.

Special care needs to be taken when computing the scene fitting error $\epsilon_{\mathbf{X}}$ in (6) as there are two cylinders C_r, C_b in the T-intersection model. Specifically, we need to assign each scene point to one of the cylinders, and compute the individual error terms in (6) with respect to this cylinder. This cylinder assignments is performed by finding the distance to each of the cylinder surfaces, and selecting the cylinder for which the absolute distance is a minimum. Figure 4c shows the results for one of the T-intersections after SBA has converged. The color-coding of the scene points (dots) represent their cylinder assignment.

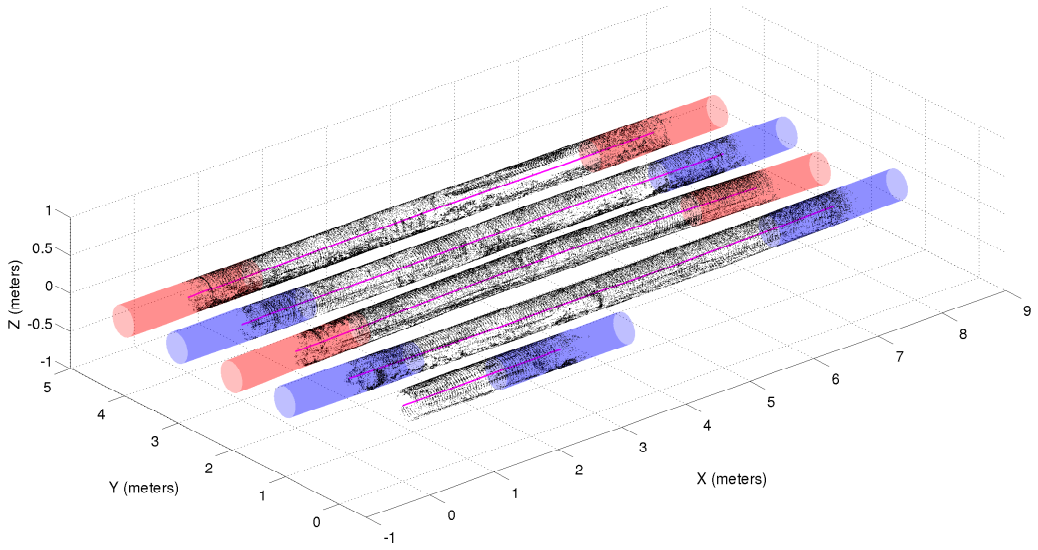
4 EXPERIMENTS AND RESULTS

A single loop datasets was collected in our constructed 400mm (16 inch) internal diameter fiberglass pipe network. The total distance traversed during the rectangular shaped loop was approximately 34 meters. The robot was tele-operating using images streamed over a wireless link, and all lighting was provided by 8 high intensity LEDs equipped on the robot — see figure 1. Over 24,000 grayscale images with $1280pix \times 960pix$ resolution were logged to the robot computer at 7.5 frames per second, from which 2760 keyframes were automatically selected. All image processing steps described in the previous section were implemented offline.

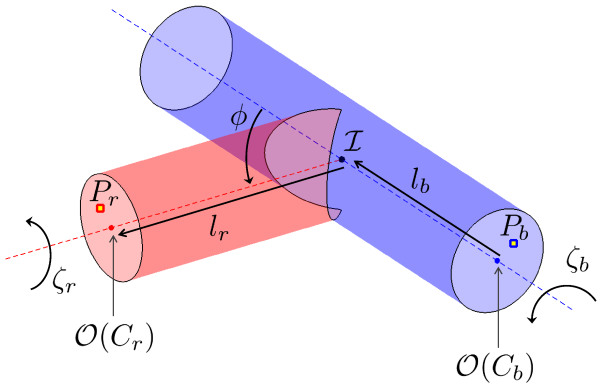
4.1 Visual Odometry

The visual odometry and sparse scene reconstruction results for each of the straight sections was show previously in figure 4a. The complete results for the pipe network are provided in figure 5. The robots start and end locations are labeled, as well as the T-intersections T1 through T4. Note that no loop closure has been used.

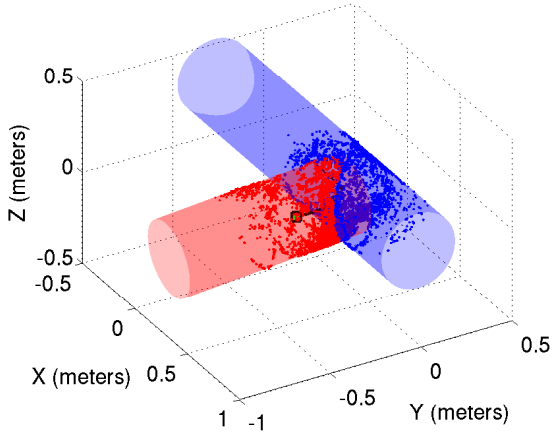
An ideal performance metric for our system is direct comparison of the visual odometry (camera pose) and scene structure results to accurate ground truth. However, obtaining suitable ground truth is particularly challenging due to the unavailability of GPS in a pipe, and limited accuracy of standard grade Inertial Measurement Units (IMUs).



(a) Straight sections with cylinders fitted to endpoints.



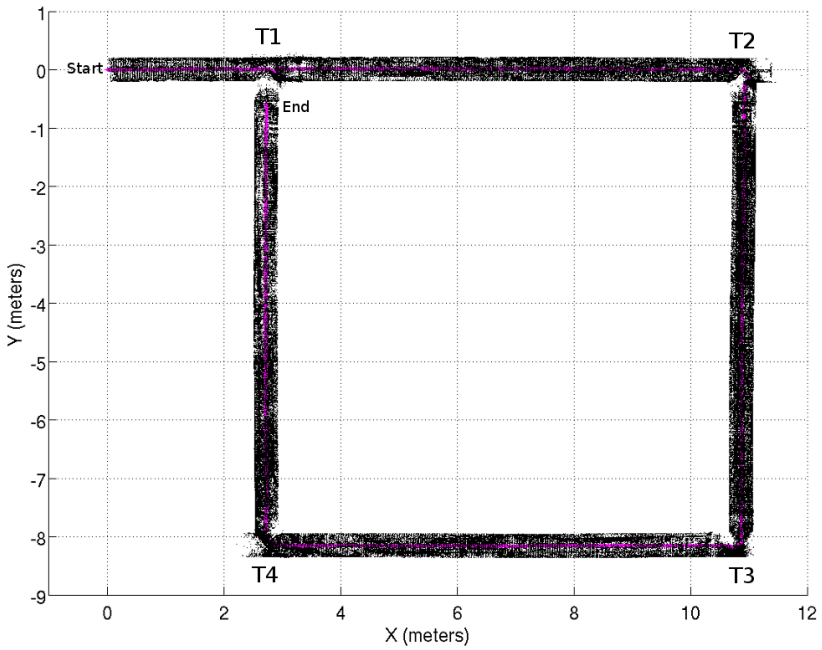
(b) The two cylinder T-intersection model parameters (refer to text for detailed description). The red cylinder is constrained to intersect the blue cylinder at a unique point $\mathcal{I}$. Both pipes are assumed to have the same internal radius r.



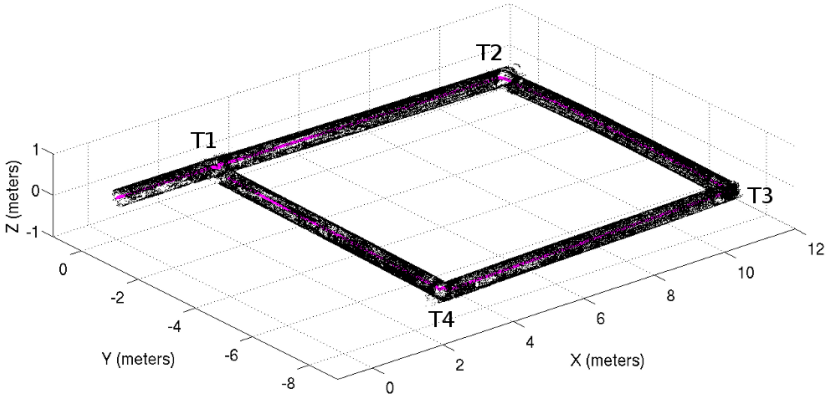
(c) Visual odometry and scene reconstruction result using the T-intersection model. The scene points have been automatically assigned to each section allowing cylinder fit regularization terms to be used within the SBA framework.

Figure 4: A T-intersection is modeled as the intersection of cylinders fitted to the straight sections of the pipe. Respectively, the blue and red colors distinguish the horizontal and vertical sections of the 'T', as illustrated in (b).

Our current ground truth is hand-held laser distance measurement estimates between all T-intersection centroids. For practical reasons, we used the center of the upper exterior surface of each T-intersection as the centroid. These measurements are compared to the distances obtained from the visual odometry in table 1. For the visual odometry, each point of intersection $\mathcal{I}$ (see figure 4c) was used as the centroid. Unfortunately, there is no means for associating the laser and visual odometry reference points used for the T-intersection centroids, and as such the ground truth measurements contain some degree of uncertainty. The largest absolute error is between T4-T1, having a value of 0.051m (0.63%). This degree of accuracy is approaching the estimated uncertainty of the ground truth measurements. We are exploring alternate techniques for obtaining more accurate ground truth.



(a) Visual odometry and sparse scene reconstruction: viewpoint 1.



(b) Visual odometry and sparse scene reconstruction: viewpoint 2.

(c) The pipe network with labeled T-intersections for reference. Observe that each of the long straight sections is constructed from two straight segments of pipe.

Figure 5: Visual odometry (magenta line) and sparse scene reconstruction (black points) for the single loop pipe network dataset.

Distance	T1-T2	T2-T3	T3-T4	T4-T1	T1-T3	T2-T4
Laser (m)	8.150	8.174	8.159	8.110	11.468	11.493
VO (m)	8.184	8.131	8.138	8.161	11.514	11.543
Error (m)	0.034	-0.043	-0.021	0.051	0.046	0.050
Error (%)	0.42	-0.53	-0.26	0.63	0.41	0.44

Table 1: The T-intersection center-to-center distances obtain with a hand-held laser (ground truth), and visual odometry – refer to figure 5. The signed error percentages are given with respect to the laser measurements.

In practice, modeling each long straight section of our pipe network as a perfect straight cylinder is too restrictive. Firstly, each individual pipe segment contains some degree of curvature/sag. Secondly, the individual segments used to construct the long straight sections of pipe (see figure 5c) are not precisely aligned. It is for this reason that we only perform straight cylinder fitting locally as part of of the 100 keyframe sliding window SBA. Doing so permits some gradual pipe curvature to be represented in the results, as evident in 4c. However, for severe pipe misalignments or elbow joints, we expect the accuracy of the results to rapidly deteriorate. Some form of cubic spline modeling of the cylinder axis may be required in these scenarios, despite the significant increase in computational expense when computing the scene point regularization errors. We aim to address these issues in future work.

A number of other directions for future work will be pursued to improve the accuracy and flexibility of the system. They include structured lighting options to estimate the internal pipe radius directly, and loop closure detection/correction (e.g. [2]).

4.2 Dense Rendering

Using the visual odometry results, an appearance map of the pipe network was produced which may be used as input for automated corrosion algorithms and direct visualization in rendering engines. Figure 6 shows the appearance map of the pipe network, including and zoomed in view of a small straight section (figure 6a) and T-intersection (figure 6b) to highlight the detail. The consistency of the appearance, namely the lettering on the pipes, demonstrates accurate visual odometry estimates.

The appearance map produced is a dense 3D grayscale point cloud which could be extended to a full facet model. The Euclidean point cloud coordinates were set using the cylinder fitting results for both the straight sections and T-intersections. The grayscale intensity value for each point was obtained by projecting the point into all valid fisheye images and taking the average sampled image value. Here, valid is defined as having a projected angle of colatitude $45^\circ < \theta < 90^\circ$ (i.e. near the periphery of the fisheye image where spatial resolution is a maximum). A range of improvements are being developed to mitigate the strong lighting variations in the rendered model. As evident in figure 6c, these are most prevalent in the T-intersections were the raw imagery can contain strong specular reflections and saturation.

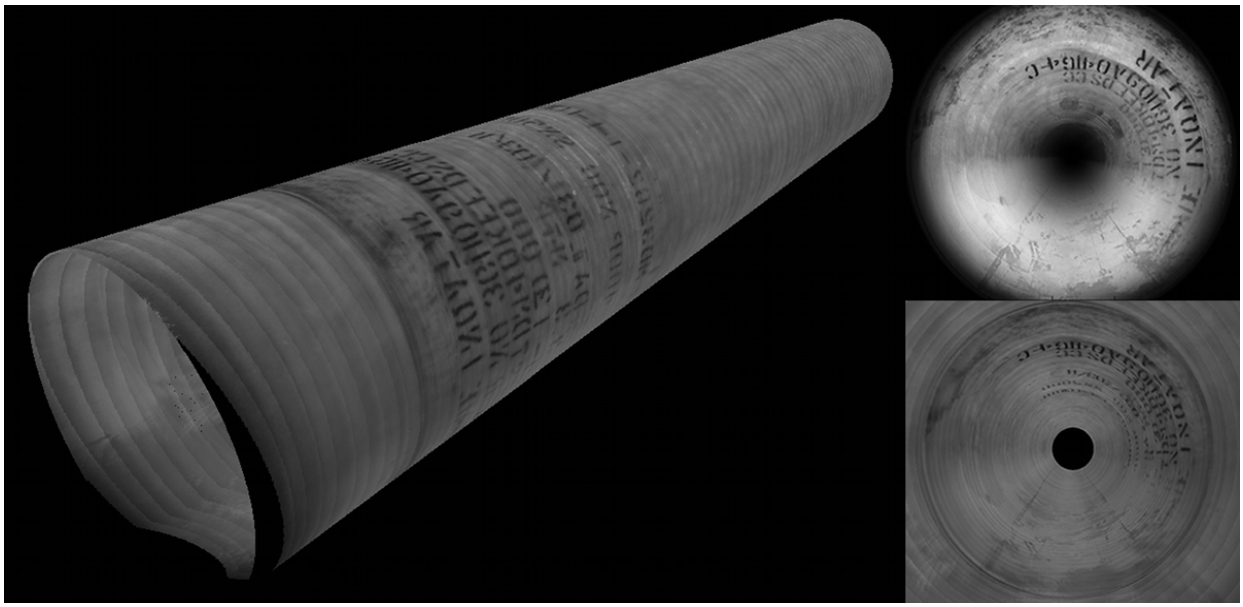
5 CONCLUSIONS

An initial fisheye visual odometry and mapping system was presented for non-destructive automated corrosion detection in LNG pipes. To improve the accuracy of the system, various cylinder fitting constraints for straight sections and T-intersections were incorporated as regularization terms in sparse bundle adjustment frameworks. The camera pose estimate and fitted cylinders are used as the basis for constructing dense pipe renderings which may be used for direct visualization.

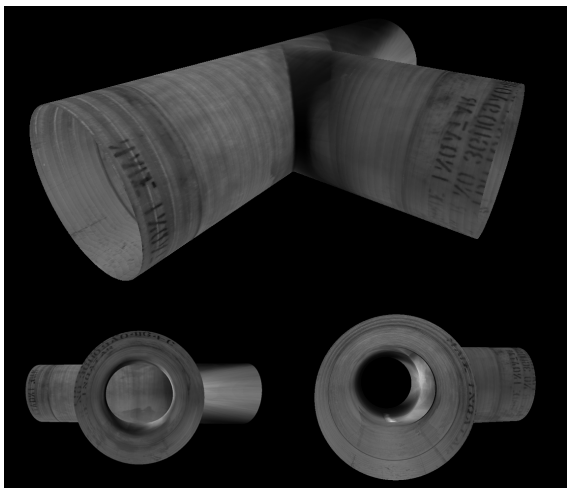
Results were presented for a single loop dataset logged in a 400mm internal diameter fiberglass pipe net-

work. To evaluate the accuracy of the pipe network reconstruction, we compared the the distances between all T-intersections. All distance measurements obtained were well within $\pm 1\%$ of the ground truth laser distance measurements. Moreover, the dense appearance maps produced further highlighted the consistency of the camera pose and scene reconstruction.

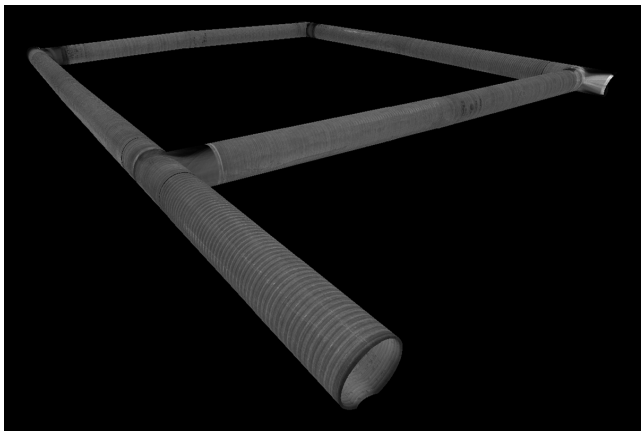
A number of directions for future work were proposed, including structured lighting for scale recovery, spline based cylinder axis modeling, and loop closure.



(a) Small straight section approximately 2 meters in length. Top right is an original fisheye image, and bottom right the reconstruction near the same location in the pipe.



(b) T-intersection.



(c) Full dataset.

Figure 6: Dense 3D grayscale appearance map of the pipe network. (a) and (b) are zoomed in sections of (c).

References

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.

[2] G Dubbelman, P Hansen, B Browning, and M. B Dias. Orientation only loop-closing with closed-form trajectory bending. In *IEEE International Conference on Robotics and Automation*, St. Paul, USA, May. 14-18 2012.

[3] Martin A Fischler and Robert C Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comms. of the ACM*, pages 381–395, 1981.

[4] Peter Hansen, Hatem Alismail, Brett Browning, and Peter Rander. Stereo visual odometry for pipe mapping. In *IROS*, 2011.

[5] C.G. Harris and M.J. Stephens. A combined corner and edge detector. In *Proceedings Fourth Alvey Vision Conference*, pages 147–151, 1988.

[6] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2003.

[7] David Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[8] David Nistér. An efficient solution to the five-point relative pose problem. *PAMI*, 26(6):756–770, June 2004.

[9] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[10] David Nistér, Oleg Naroditsky, and James Bergend. Visual odometry for ground vehicle applications. *JFR*, 23(1):3–20, January 2006.

[11] Hagen Schempf. Visual and nde inspection of live gas mains using a robotic explorer. *JFR*, Winter, 2009.

[12] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ICCV '99, pages 298–372, London, UK, 2000. Springer-Verlag.