3D Pipe Network Reconstruction Based on Structure from Motion with Incremental Conic Shape Detection and Cylindrical Constraint

Sho Kagami[1], Hajime Taira[1], Naoyuki Miyashita[2], Akihiko Torii[1], and Masatoshi Okutomi[1]

1 Dept. of Systems and Control Engineering, Tokyo Institute of Technology

2 R&D Group, Olympus Corporation

Abstract— Pipe inspection is a critical task for many industries and infrastructure of a city. The 3D information of a pipe can be used for revealing the deformation of the pipe surface and position of the camera during the inspection. In this paper, we propose a 3D pipe reconstruction system using sequential images captured by a monocular endoscopic camera. Our work extends a state-of-the-art incremental Structure-from-Motion (SfM) method to incorporate prior constraints given by the target shape into bundle adjustment (BA). Using this constraint, we can minimize the scale-drift that is the general problem in SfM. Moreover, our method can reconstruct a pipe network composed of multiple parts including straight pipes, elbows, and tees. In the experiments, we show that the proposed system enables more accurate and robust pipe mapping from a monocular camera in comparison with existing state-of-the-art methods.

I. INTRODUCTION

The needs for pipe inspection are rapidly increasing in various plants, such as chemical refineries, gas distribution, and sewer maintenance. Since the damage and clogging in the pipe are substantially related to disastrous failures of the whole system, periodic industrial inspection is necessary to keep its function. An industrial endoscope, or an industrial videoscope, is commonly used to inspect the inside of a pipe, since it is impossible for people to directly access the inspection site such as gas network underground or inside the building structures. When an operator inserts the probe inside the pipe and then visually inspects on a remote screen to reveal the deformations or defects, they need to guess the defect location and 3D structures around the camera only from the 2D image sequences.

Vision-based 3D reconstruction techniques such as Structure-from-Motion (SfM) [1]–[9] and Visual Simultaneous Localization and Mapping (Visual SLAM) [10]–[12] can potentially help those situation, by reconstruting the 3D structure of the pipe from images, and localizing the probe trajectory during the operation. However, because of extraordinarily repetitive and narrow structures as shown in Fig. 1, existing methods, which are mostly tested on the urban situation, fail to reconstruct the scene or produce an erroneous structure.

In this paper, we propose a incremental SfM system for a pipe network, with a monocular camera used for the inspection. To address the error arisen from the challenging appearances and narrow geometries of the industrial parts, we use prior information of the pipe (assuming a constant inner diameter). In contrast to relevant works [13], [14], our system
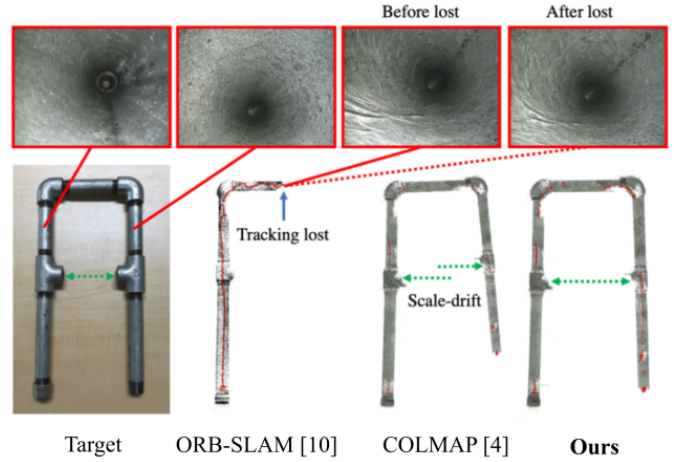

Fig. 1. **Difficulties of pipe reconstruction.** Narrow and repetitive structures inside of a pipe make vision-based 3D reconstruction extremely difficult. The reconstruction from such difficult conditions leads to severe drift errors and failure of tracking. We make a SfM system tailored to the pipe structure.

is carefully designed to incorporate the prior constraint into the iterative process of SfM, which enables the system to deal with large errors of the reconstruction and to be capable for the pipe network consisting of multiple straight pipes. During the reconstruction, a temporary 3D model is incrementally updated regarding feature correspondences between the previous model and the current image frame from the camera. Considering the potential errors induced by scale-drifting, we find multiple straight pipes in the temporary model as conic shapes of the structure. Subsequently, our proposed cylindrical constrained BA incrementally refines each pipe to be in line with the given prior. Experiments on practical pipe networks consisting of multiple different parts clearly show that our methods obtain more accurate reconstruction compared with the state-of-the-art vision-based methods.

II. RELATED WORK

A. Vision-based 3D reconstruction

For a set of images input, SfM depicts the captured scene by 3D scene points, based on local feature matching [15], [16] and multi-view triangulation [17]–[19], while estimating the camera poses for the input images. In the recent decade, a variety of SfM strategies, including incremental [1]–[4], hierarchical [5], global [6]–[8], and hybrid approach of

them [9] have been studied. For a sequential image series input, incremental SfM is the most popular strategy that can be extended to the real-time application [20]. On the other hand, SLAM-based methods have been developed in contexts of the real-time operation of estimating the camera trajectory while reconstructing the environment. ORB-SLAM [10] speeds up the feature extraction process by using binarized ORB descriptor [21]. Direct methods [11], [12] allow more efficient operation directly obtaining camera trajectory by minimizing a cost regarding the differences of image intensities.

One common issue of those vision-based reconstruction methods is the accumulated scale-drift problem that causes a camera trajectory and a 3D model inaccurate. Several works address the issue via loop closure [22]–[24] that attempts to detect a camera path loop, *e.g.*, using image appearance [25], [26], and hence detect the scale-drift during the reconstruction. The 3D points and camera poses of the model are then refined as the model keeps consistent scene geometries at the former and the latter of the camera path loop. Another approach is to compensate scale information via pre-trained deep architectures that estimate an absolute scale depth [27], [28] and/or relative motion between the frames [29]–[32].

B. Pipe reconstruction

Pipe network for gas distribution or sewer is an active target of 3D reconstruction since they are often narrow or dangerous to walk in for inspection. Instead, a robot vehicle [13] or an industrial endoscope [33], [34] provides a safety inspection using its mounted camera(s). 3D reconstruction from a sequence of images from the camera could also help a detailed 3D structure inspection rather than the 2D visual inspection. However, poor and dynamically changing lighting conditions and highly repetitive appearances due to standardized pipes make application of vision-based approaches difficult. Several works, therefore, rely on the use of multi-domain sensors outputs, *e.g.*, stereo camera [35], Inertial Measurement Unit (IMU) [36] or structured light [13], along with a fish-eye camera. Although rich sensors and equipment can help reconstruction, they are sometimes infeasible for very narrow structures and limited conditions, *e.g.*, inspection of gas plumbing inside the building using an endoscope. Some works build an accurate 3D pipe model using only a monocular camera, with assumptions of camera motion and prior knowledge peculiar to the pipe. Kahi *et al.* [14] refine the reconstructed 3D points by cylinder fitting after BA. Zhang *et al.* [37] and Kunzel *et al.* [38] rectify the images to a cylinder projection plane to triangulate points easily and regularize image illumination.

Our system, designed for a 3D reconstruction using an industrial endoscope equipping only a monocular camera, is built based on the incremental SfM pipeline. Assuming a prior knowledge of the constant inner diameter of the straight pipe, we provide an accurate 3D structure by fitting 3D points to the known pipe surface. In contrast to the previous works limiting a camera movement to be in parallel with the pipe axis [37], [38], we give no limitations about the camera path, which enables general endoscope motions during the inspection. The most relevant work of ours is [14], which detects a cylinder per reconstructed model and aligns the 3D points to the known pipe property after registering the fixed number of frames. On the other hand, we detect the pipe as a general conic shape, which takes the scale-drifting errors of 3D points into account. More importantly, we search multiple pipe instances assuming the pipe network that consists of multiple pipes directed to different axes. We then incrementally refine the temporary model using a known inner diameter when each of new pipes is appeared. The next section describes our SfM system and its components in detail.

III. CYLINDER CONSTRAINED SfM FOR PIPE RECONSTRUCTION

In this section, we describe the proposed SfM system for reconstructing a pipe network composed of multiple straight pipes using an image sequence taken by an endoscope camera. The main challenges in the situation are three-folded; **1)** Pipe networks constructed for industrial purposes are dominated by very weakly-textured (or highly repetitive) appearance (cf. Fig. 4), and often consists of narrow and specular reflective cylindrical parts. In addition, pipe inspection using a endoscopic camera is often operated in a poor lighting condition, consequently suffers from local feature deprivation and inaccurate keypoints. Vision-based system such as SfM and SLAM are critically affected by such unstable keypoint properties, resulting in an inconsistent 3D reconstruction. **2)** Incremental SfM iteratively registers each image in the image sequence. Therefore, the error of the temporary 3D model in each iteration is increasingly accumulated to the whole model. **3)** Loop closure is often infeasible during the practical visual inspection, because the camera path often has no loops due to the flexibility of the endoscope, *i.e.*, the system cannot detect the scale-drift.

We construct our reconstruction system (illustrated in Fig. 2) based on the incremental SfM pipeline for general purposes (Sec. III-A), while integrating several configuration and new processes to achieve an accurate reconstruction for the pipe network: Before starting reconstruction, we calibrate intrinsic parameters of the endoscope camera to mitigate the effect of lens distortion (Sec. III-B). Our system finds each pipe instance from the temporary reconstructed model considering its scale-drifting (Sec. III-C). Detected pipes are refined by our cylinder constrained BA that minimizes point distances from the cylinder surface using a known property of the tube diameter (Sec. III-D).

A. Incremental SfM

In what follows, we describe an general incremental SfM pipeline for a set of sequential images.

2D feature matching. For each input image, SfM extracts the image local features, *e.g.*, SIFT [15] or improved one [16], to get 2D correspondences between images that are used to register the image to the model in a latter step. When the set of images are ordered by time-stamp, the matching target can be restricted within the current few frames. In our experiment,
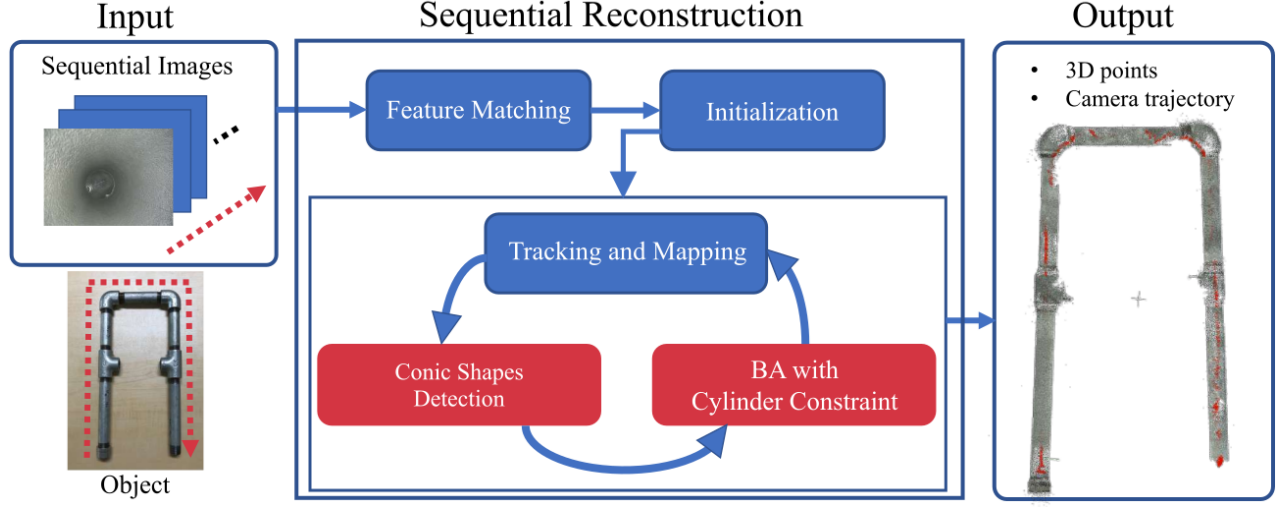
Fig. 2. **The overview of our SfM system for pipe reconstruction.** We design our system based on the incremental SfM pipeline for general purposes while extending it to address the particular situations of the inner pipes. Our system detects multiple pipe instances as general conic shapes in the temporary model (conic shape detection) and refines the whole model as each of the detected pipes satisfies the known inner diameter (BA with cylinder constraint).

we match an image toward the current 50 frames of the sequential set. Feature correspondences are verified through an outlier rejection scheme, *e.g.*, random sample consensus (RANSAC) [17], [39].

3D model initialization. Incremental SfM firstly initializes the 3D model for a selected image pair [4], [40]. We assume the input from a sequential image set, thus the selection can be easily done using time stamp, *i.e.*, initializing the model with first two frames of the input. Then the initial model, composed of the 3D scene points correspond to the local feature matches and relative camera poses of the image pair, is to be constructed via two-view triangulation [19], [40].

Tracking and mapping (temporary model construction). Once the model has been initialized, the SfM incrementally registers the input images to the model, while enriching the model by adding new 3D points correspond to 2D local feature tracks. In each iteration, SfM obtains local feature correspondences for a new input image (next frame) to the existing model, resulting in the set of tracked 2D observations, *i.e.*, the keypoints seen from more than two other frames.

Using the existing 3D points correspond to the feature tracks, SfM estimates the camera pose of the input image by solving a Perspective-n-Point (PnP) problem via P3P-RANSAC [39], [41]. After recovering the pose of the image, the model grows adding 3D points for the newly tracked features, through the triangulation among the current frames.

Bundle adjustment. To stably develop the model through the incremental scheme, the system refines the temporary 3D model after each input image registration (local bundle adjustment). Regarding the 3D points and camera poses of the current frames, the standard bundle adjustment [42] minimizes the error of the 3D points from the corresponding 2D

observations (reprojection error), which is represented as:

$$E_{rep}(\mathbf{X}, \mathbf{P}, \mathbf{K}) = \sum_{i \in \mathbf{P}} \sum_{j \in \mathbf{X}} \rho(\|q_{ij} - \pi(\mathbf{P}_i, \mathbf{K}_i, \mathbf{X}_j)\|) \quad (1)$$

where $\mathbf{X}_j$ is the j–th 3D scene point, q_{ij} is the 2D observation of $\mathbf{X}_j$ from the i–th view $\mathbf{P}_i$, $\pi(\mathbf{P}_i, \mathbf{K}_i, \mathbf{X}_j)$ is a function that projects scene points to the image plane, and ρ is the robust function, *e.g.*, Cauchy function.

The system also runs another refinement process after several iterations (global bundle adjustment) that maintains the consistency of the whole model. In this time, bundle adjustment also minimizes Eq. (1) but in regard to all 3D points and frames registered to the model.

B. Endoscope camera calibration

An accurate intrinsic parameter of the camera, that makes a relation between an image point $[u, v]^{\mathrm{T}}$ and a 3D point (normalized image coordinate) $[x, y, z]^{\mathrm{T}}$, is required to achieve a solid 3D reconstruction, *e.g.*, for an accurate conversion π in Eq. (1). We focus on the use of a standard industrial endoscope (Fig. 3, Tab. I) which has a wide FoV camera for efficient visual inspection. To deal with the image distortion comes from the wide FoV configuration, we rectify the image coordinates of the keypoints assuming a fish-eye model [43]. By the camera model, the relation between the image coordinate and the 3D point is formulated as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x(\theta + k_1\theta^3 + k_2\theta^5)cos\phi + u_0 \\ f_y(\theta + k_1\theta^3 + k_2\theta^5)sin\phi + v_0 \end{bmatrix} \quad (2)$$

where θ is the 3D angle formed by the camera optical axis and the ray going from camera center to the 3D point, and ϕ is the polar angle of normalized image coordinate, respectively.

$$\theta = \arctan\left(\sqrt{\frac{x^2 + y^2}{z^2}}\right), \phi = \arctan\left(\frac{y}{x}\right) \quad (3)$$
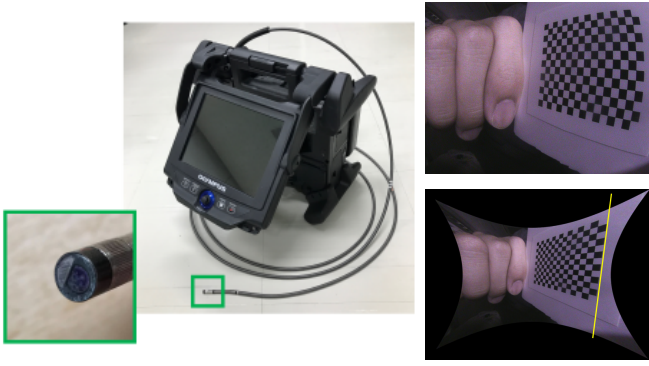
Fig. 3. **Industrial endoscope.** Left: The appearance of the industrial endoscope. We use an Olympus IPLEX NX with the AT120D/NF-IV96N optical adapter and the IV9635N scope. Right: A sample image capturing checkerboard pattern (top) and rectified by fish-eye camera model (bottom).

Table I
THE DETAILED SPECIFICATION OF THE INDUSTRIAL ENDOSCOPE

Resolution	Field of view	Scope diameter	Depth of field	Illumination
1024×768[px]	$120°$	6.0 mm	7 to 300 mm	laser diode

Camera intrinsic parameter consists of: focal length with respect to horizontal and vertical axis (f_x, f_y), distortion parameters (k_1, k_2), and image principal point (u_0, v_0). Before starting the reconstruction, we initialize the parameters by an offline calibration. To achieve an accurate calibration, we use a checkerboard pattern with a 2mm size of each square (Fig. 3) and take pictures from multiple views. The parameters are found by minimizing the sum of squared reprojection errors of the grid points [44]. We also update the parameters during the reconstruction via bundle adjustment (Sec. III-D).

C. Incremental conic shapes detection

In spite of the effort of the precise camera calibration, general incremental SfM pipeline can often cause a significant 3D model distortion on a pipe inner situation. We address the errors assuming the known properties of a pipe network, *i.e.*, if we can detect the pipe instances that is a component of the pipe network, the model can be refined by fitting 3D points to the known properties. Instead of detecting the cylinders for the whole 3D model, which would requires substantial conversion of the model because of accumulated model distortion, we search the pipe instance from the temporary 3D points during the incremental pipeline.

Also, we observed even a small intrinsic calibration error can cause a large scale-drifting of the temporary model (Fig. 5), which can limit the standard cylinder detection for the 3D point cloud. Therefore, we search the pipe by fitting a general conic shape [45] to the 3D points.

In each temporary model constructed during the incremental SfM process, we assume a 3D point in the homogeneous coordinate system $\boldsymbol{X} = [x, y, z, 1]^{\mathrm{T}}$ represents the surface of a cone if it satisfies:

$$\boldsymbol{X}^{\mathrm{T}} \mathbf{C} \boldsymbol{X} = 0 \qquad (4)$$

$\mathbf{C}$ is a symmetric matrix, which can be decomposed as:

$$\mathbf{C} = \begin{bmatrix} R^{\mathrm{T}} \mathbf{D} R & R^{\mathrm{T}} \mathbf{D} t \\ t^{\mathrm{T}} \mathbf{D} R & t^{\mathrm{T}} \mathbf{D} t \end{bmatrix}, \mathbf{D} = \mathrm{diag}(-c^2, -c^2, 1) \qquad (5)$$

where c is the constant parameter representing the slope of the cone, and $[R, t]$ is the 3D rotation and translation that represents a coordinate transformation which aligns z–axis to the major axis of the cone. Eq. (4) can also be rewritten by:

$$vech^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}) vech(\mathbf{C}) = 0 \qquad (6)$$

$vech()$ is the half vectorization transformation of a symmetric matrix that is obtained by vectorizing the lower triangular part of the matrix. Minimal solution for Eq. (6) is therefore given by nine points.

We incrementally find multiple pipe instances by fitting cone to the newly produced 3D points. In each iteration of the incremental SfM, we search a cone in the temporary 3D model via RANSAC [39]. The registered images are then labeled as it belongs to the pipe or not, based on the number and the ratio of inliers that support the cone model. Once a pipe instance is detected, the detector searches a new conic shape from the 3D points seen by current frames that do not belong to any existing instances. All cone parameters are then refined by local hypotheses refinement [46] using 3D points observed by the labeled images, which achieves optimal cone fitting for 3D points.

D. Bundle adjustment with cylinder constraint

Bundle adjustment in the general incremental SfM pipeline refines the current model by minimizing reprojection error represented by Eq. (1). If the conic shape detector finds the straight pipes, we can also compute the error of the 3D points with respect to the prior knowledge of the pipe properties, which is formulated as:

$$E(\mathbf{X}, \mathbf{P}, \mathbf{K}, \mathbf{C}) = E_{rep}(\mathbf{X}, \mathbf{P}, \mathbf{K}) + \alpha E_{cyl}(\mathbf{X}, \mathbf{C}) \qquad (7)$$

where the first term $E_{rep}(\mathbf{X}, \mathbf{P}, \mathbf{K})$ is the reprojection error term which is equal to Eq. (1), and the second term $E_{cyl}(\mathbf{X}, \mathbf{C})$ is our new cylinder constraint term. $\mathbf{X}, \mathbf{P}, \mathbf{K}$, and $\mathbf{C}$ are the variables that indicate the sets of 3D points, the camera poses of the registered images, the camera intrinsic parameters, and the detected cones parameters, respectively. α is a constant scalar which controls the weights of two competing error terms.

Our cylindrical constraint punishes the distance of the 3D points $\boldsymbol{X}_j$ from the cylinder surface around the major axis. Assuming the uniformly distributed accumulated error, the axis of the cylinder can be approximated by detected axis of the cone $[R_i, t_i]$, which is given by the decomposition of cone parameters $\mathbf{C}_i$ according to Eq. (5). E_{cyl} is therefore formulated by:

$$E_{cyl}(\mathbf{X}, \mathbf{C}; r) = \sum_i \sum_j \rho(\|r - d(\boldsymbol{X}_j, [R_i, t_i])\|) \qquad (8)$$

where $d(\boldsymbol{X}_j, [R_i, t_i])$ is the distance of the 3D point $\boldsymbol{X}_j$ from the major axis of the cylinder, and r is the known inner
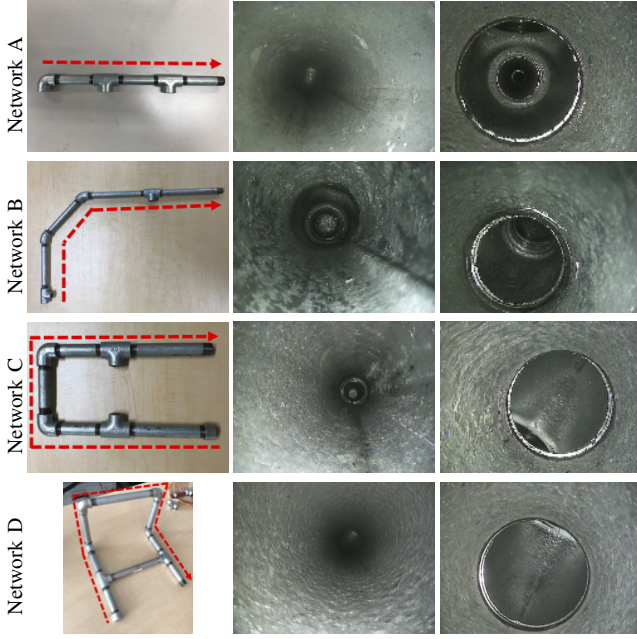
Fig. 4. **Datasets.** Left: the target pipe network and camera path (red arrow) of each sequence. Middle and right: sample images of each sequence (middle: at a straight pipe, right: at an elbow).

Table II
The properties of each dataset. WE SETUP FOUR DIFFERENT TYPES OF PIPE NETWORKS AND CAPTURE VIDEOS USING THE STANDARD INDUSTRIAL ENDOSCOPE AS SHOWN IN FIG. 3. NOTE THAT THE MATERIAL OF ALL PIPES IS STEEL.

Properties	A	B	C	D
Inner diameter [mm]	16.1	8.0	16.1	16.1
Video times [sec]	61	183	367	442
Centering device	Yes	No	No	No
# Straight pipe	3	4	5	7
# Tee	2	2	2	2
# Elbow	1	2	2	4
Elbows angle (max)	0°	45°	90°	90°

diameter of the pipe. ρ is the Cauchy function used as the robust function.

Our incremental SfM system includes two types of bundle adjustment, local BA and global BA. After registering each input image, our system performs local BA that refines camera poses of current frames, intrinsic parameters, and 3D points, by minimizing Eq. (7). When the model grows by a certain percentage or a new pipe instance is detected, the system runs global BA that optimizes all model parameters including the straight pipe parameters. For completeness and fastness, we refine cameras intrinsic only after detecting first straight pipe. Please notice that our newly proposed error term does not give any constraint on camera motion, unlike the previous works [14], [37], [38]. Also notice that our SfM system updates each temporary model, thus produces a substantial different results from the one constructed via standard SfM at the end. As shown in the next section, this property enables us to obtain further complete and accurate 3D model (cf. Fig. 6).

IV. EXPERIMENTS

In this section, we describe the performance of our incremental SfM pipeline designed for a pipe network reconstruction. We evaluate our method on several image sequences capturing the inners of pipe networks, which consist of multiple industrial pipes and indicate the practical scenarios of industrial visual inspection. We firstly demonstrate our new constrained BA in an incremental SfM system effectively deals with error accumulation during the reconstruction (Sec. IV-A). We next compare our method with several vision-based reconstruction systems on our dataset (Sec. IV-B).

Environment. Fig. 4 shows the four types of pipe networks we set up for evaluation. Detailed properties are described in Tab. II. All scenes consist of 1∼5 straight pipes, tees, and elbows, constructed by industrial steel parts. We collect 60fps image sequences of those networks using an industrial endoscope (Fig. 3). Detailed specifications of the endoscope are summarized in Tab. I. The camera moves backward inside the networks following the red arrows in Fig. 4, which is the practical manner of the pipe inspection due to the physical limitation of the endoscope.

The network A consists of three straight pipes in the same direction. In this simple structure, we optionally attach an guide head device for the endoscope that roughly forces the camera to be in center of the pipe. Note that this setting fairly makes the camera path to be stable and makes the reconstruction easier, but is sometimes infeasible, because the guide head does not support significant direction changes, *e.g.*, curved camera path at an elbow. We do not use this attachment for network B, C, and D, to depict more general situations of the pipe network inspection. Network B has a narrower inner diameter (8.0mm) than others (16.1mm) which leads to a more severe appearance changes and occlusions at the elbow parts. Network D consists of the maximum number of pipes and elbows according to the flexibility of the endoscope, connected three-dimensionally with independent orientation.

Evaluation metric. To evaluate the accuracy of 3D models, we compute the reconstruction error as the difference from the prescribed inner diameter of each straight pipe. RMSE of radius rate is determined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(\frac{r_i - r}{r}\right)^2} \qquad (9)$$

where r_i is the radius of inlier points in the straight cylinders estimated by our SfM, and r is the prescribed radius value. For other methods that originally do not detect any pipe, we additionally detect cylindrical parts and scale the model for evaluation, after the whole reconstruction process. Specifically, we fit the multiple cones to the reconstructed model via sequential RANSAC while giving the number of pipe parts. The model is then scaled as approximating the diameter of the pipe by the average of points distances from the cone axis.

Implementation. We construct our system based on an incremental SfM system implemented by COLMAP [4], a widely

Table III
Quantitative results. EVALUATION OF 3D RECONSTRUCTION RESULTS BY OUR DATASETS. THE VALUES ARE THE RADIUS ERROR RATE (RMSE) (EQ. (9)).

Method	A	B	C	D
ORB-SLAM	0.1916	0.5428	0.3331	**0.0958**
COLMAP	0.1615	0.3291	0.3055	0.2670
Ours	**0.1375**	**0.2719**	**0.1560**	0.1034

known reconstruction tool. After constructing temporary 3D model for each 30 frames of the input, the system searches and refines the pipe instances of the model as described in Sec. III-C. Once a cylinder is detected, the system replace bundle adjustment process in each iteration by our cylinder constrained BA (Sec. III-D). We assume the pipe inner diameter of each Network is constant and known (cf. Tab. II). We experimentally set the parameter α in Eq. (7) by 10.

A. Impact of cylinder constrained BA

To demonstrate the impact of our cylinder constrained BA using the known pipe property, we evaluate our method on Network A, the simplest situation on which all pipes direct to a common axis. For a comparison, we also run an incremental SfM for general purposes (COLMAP) [4], which refines the model by minimizing Eq. (1), on the same sequence. Fig. 6 (c,d) shows the appearances of the models obtained by proposed SfM and COLMAP. Both methods succeed to recover the whole pipe network but proposed provides more accurate model (cf. Tab. III). Fig. 5 shows the 3D points of the temporary models obtained during the reconstruction, via COLMAP (a) and proposed (b). During the reconstruction, COLMAP increasingly produces large errors regarding the known inner diameter, due to the errors of intrinsic parameters and erroneous estimations of the camera motion. On the other hand, proposed method iteratively detects and refines each of the pipe instances, resulting in more accurate 3D model regarding the constant inner diameter.

B. Reconstruction of multiple pipes network

Next we compare our SfM system with several other reconstruction systems on the pipe networks consist of multiple straight pipes.

Comparisons. We compare our method to the state-of-the-art methods in each of three approaches described in Sec. II, COLMAP [4] for SfM, ORB-SLAM [10] for feature-based SLAM, and DSO [11] for direct SLAM, respectively. For each comparison, we use the implementation provided by the authors. For a fair comparison, we use a calibrated fish-eye camera model (Sec. III-B) for all methods[1]. Note that we use DSO without photometric calibration since it is difficult to collect calibration data for industrial endoscope because of a built-in light source. While we use input sequence as 60 fps for real-time methods (ORB-SLAM and DSO), we use 5 fps sequence for offline methods (ours and COLMAP). We do

---

[1]Since the original ORB-SLAM implementation only accepts a perspective model, we extend it for the fish-eye model.

(a) 3D points obtained via COLMAP.

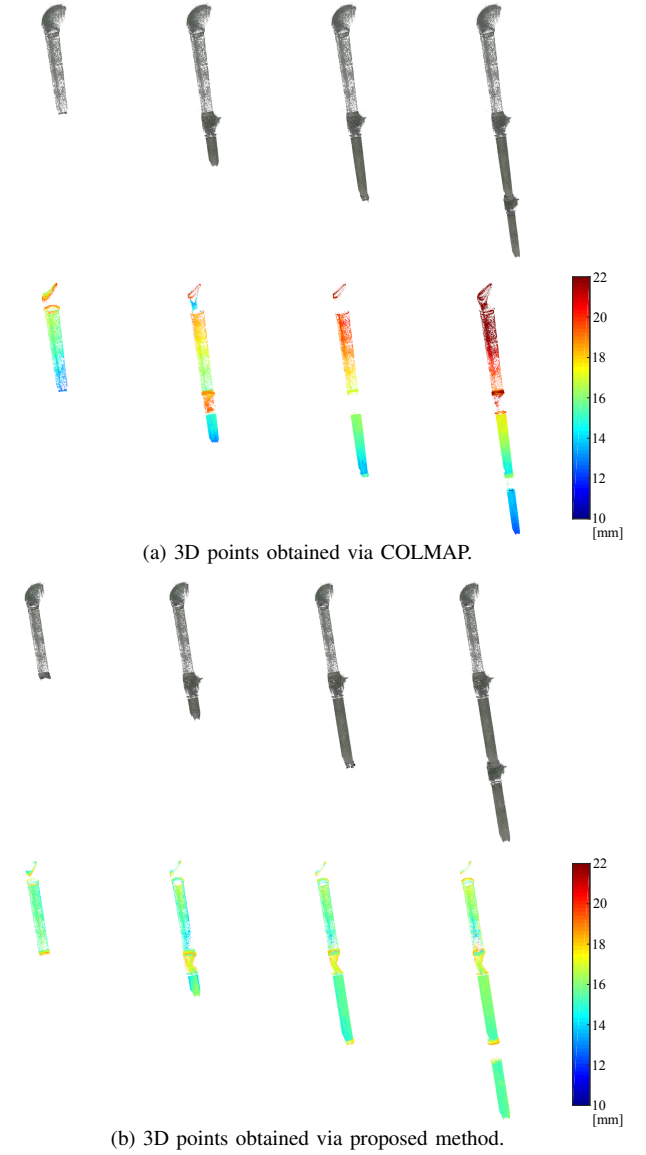(b) 3D points obtained via proposed method.

Fig. 5. **Progress of the temporary model during the incremental SfM reconstruction.** 3D points show how the temporary model obtained via (a) COLMAP and (b) our incremental SfM pipeline grow up when each of the new images is registered. From left to right, each column roughly associates the temporary model updated by the end of the first pipe, the middle of the second pipe, the end of the second pipe, and the end of the pipe network, respectively. Color-coded 3D points indicate the distance of each 3D point from the major axis of the pipe, regarding the true diameter of the pipes as 16.1mm.

not compare our methods with several works designed for a single pipe [13], [14], [37], [38] because they do not provide their original implementations. But we believe adapting these works for each reconstructed pipe as batch-like process does not much improve reconstruction since they highly depend on the quality of the initial model, which is often largely distorted as shown in Sec. IV-A.

Tab. III shows the quantitative evaluation of 3D reconstruction results of COLMAP [4], ORB-SLAM [10], and ours. Our method outperforms the baseline COLMAP on all scenes, and

the margin is remarkable in network C. For network D, ORB-SLAM gives the best RMSE score, but it reconstructs only two pipes in the scene. Fig. 6 shows the qualitative results of each method in each scene. While ORB-SLAM and DSO can reconstruct in real-time, their reconstruction results are not as accurate as offline methods like ours and COLMAP, especially in complex scenes as network C and D. ORB-SLAM reconstructs all the images in A and B, but fail to track the image sequences in C and D. This failure of tracking occurs because the method assumes a constant velocity motion to track the camera that is often not functional for an endoscopic cameras motion, *e.g.*, significant view changes in an elbow. In contrast, the proposed method reconstructs the whole target for all sequences, while preserving a stable diameter.

Limitation. To address the difficulties raised in the pipe inner situation, our system applies a prior knowledge (*i.e.*constant inner diameter) to the existing 3D points in the model, rather applying before or during the 3D mapping, *e.g.*, feature matching guided by known properties, or 3D points triangulation constrained within the known pipe surface. This strategy, however, could result in insufficient model reconstruction when the system cannot find sufficient matches. The problem is caused when the pipe network includes pipes which have especially severe properties, *e.g.*, material which has a smooth nature. Potential approach to make the system robust to such severe conditions is that obtaining matches in dense manner [47], [48] attempting to get pixel-wise precise matches and offering outlier rejection scheme guided by pipe properties.

Another future work is to determine a proper parameter α in Eq. 7, which balances the temporal property and the prior information, also regarding the demand of the model quality.

V. Conclusion

In this paper, we have proposed a vision-based pipe reconstruction system that can provide an accurate 3D reconstruction for industrial endoscopic images. To deal with the accumulated model errors, our method incorporates the prior information of the pipe network without limiting the flexibility of camera motion. Proposed SfM pipeline consists of robust pipe detection and bundle adjustment constrained by the geometrical properties of a pipe system, which are carefully combined into an incremental image registration process, for stable camera tracking and 3D reconstruction. Throughout the experiments on realistic pipe network environments, it is demonstrated that our method can suppress scale drifting and reconstruct 3D pipe models more accurately and robustly than existing state-of-the-art methods. One of the future works is to develop a real-time application for giving instant feedback to the inspector.

References

[1] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," in *ACM Trans. Graphics*, vol. 25, no. 3. ACM, 2006, pp. 835–846.

[2] J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, *et al.*, "Building Rome on a cloudless day," in *Proc. ECCV*, 2010.

[3] C. Wu, "Towards linear-time incremental structure from motion," in *Proc. 3DV*, 2013.

[4] J. L. Schonberger and J.-M. Frahm, "Structure-from-Motion revisited," in *Proc. CVPR*, 2016.

[5] R. Gherardi, M. Farenzena, and A. Fusiello, "Improving the efficiency of hierarchical Structure-and-Motion," in *Proc. CVPR*, 2010.

[6] D. Crandall, A. Owens, N. Snavely, and D. Huttenlocher, "Discrete-continuous optimization for large-scale structure from motion," in *Proc. CVPR*, 2011.

[7] K. Wilson and N. Snavely, "Robust global translations with 1DSfM," in *Proc. ECCV*, 2014.

[8] C. Sweeney, T. Sattler, T. Hollerer, M. Turk, and M. Pollefeys, "Optimizing the viewing graph for Structure-from-Motion," in *Proc. ICCV*, 2015.

[9] H. Cui, X. Gao, S. Shen, and Z. Hu, "HSfM: Hybrid Structure-from-Motion," in *Proc. CVPR*, 2017.

[10] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[11] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *PAMI*, vol. 40, no. 3, pp. 611–625, 2018.

[12] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. ECCV*, 2014.

[13] P. Hansen, H. Alismail, P. Rander, and B. Browning, "Visual mapping for natural gas pipe inspection," *Intl. J. of Robotics Research*, vol. 34, no. 4-5, pp. 532–558, 2015.

[14] S. El Kahi, D. Asmar, A. Fakih, J. Nieto, and E. Nebot, "A vison-based system for mapping the inside of a pipe," in *Proc. ROBIO*, 2011.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[16] J. Dong and S. Soatto, "Domain-size pooling in local descriptors: DSP-SIFT," in *Proc. CVPR*, 2015.

[17] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge university press, 2003.

[18] R. I. Hartley and P. Sturm, "Triangulation," *CVIU*, vol. 68, no. 2, pp. 146–157, 1997.

[19] N. Snavely, "Scene reconstruction and visualization from internet photo collections," Ph.D. dissertation, University of Washington, 2008.

[20] S. Bronte, M. Paladini, L. M. Bergasa, L. Agapito, and R. Arroyo, "Real-time sequential model-based non-rigid SFM," in *Proc. IEEE/RSJ Conf. on Intelligent Robots and Systems*, 2014.

[21] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. ICCV*, 2011.

[22] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *IEEE Trans. Robotics*, vol. 24, no. 5, pp. 1066–1077, 2008.

[23] C. Mei, G. Sibley, M. Cummins, P. M. Newman, and I. D. Reid, "A constant-time efficient stereo slam system." in *Proc. BMVC*, 2009, pp. 1–11.

[24] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Scale drift-aware large scale monocular SLAM," in *Proc. Robotics: Science and Systems*, 2010.

[25] D. Filliat, "A visual bag of words method for interactive qualitative localization and mapping," in *Proc. Intl. Conf. on Robotics and Automation*, 2007.

[26] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Trans. Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[27] N. Yang, R. Wang, J. Stückler, and D. Cremers, "Deep virtual stereo odometry: Leveraging deep depth prediction for monocular direct sparse odometry," in *Proc. ECCV*, 2018.

[28] K. Tateno, F. Tombari, I. Laina, and N. Navab, "CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction," in *Proc. CVPR*, 2017.

[29] B. Ummenhofer, H. Zhou, J. Uhrig, N. Mayer, A. Ilg, A. Dosovitskiy, and T. Brox, "DeMoN: Depth and motion network for learning monocular stereo," in *Proc. CVPR*, 2017.

[30] T. Zhou, M. Brown, N. Snavely, and D. G. Lowe, "Unsupervised learning of depth and ego-motion from video," in *Proc. CVPR*, 2017.

[31] C. Tang and P. Tan, "BA-Net: Dense bundle adjustment network," *arXiv preprint arXiv:1806.04807*, 2018.
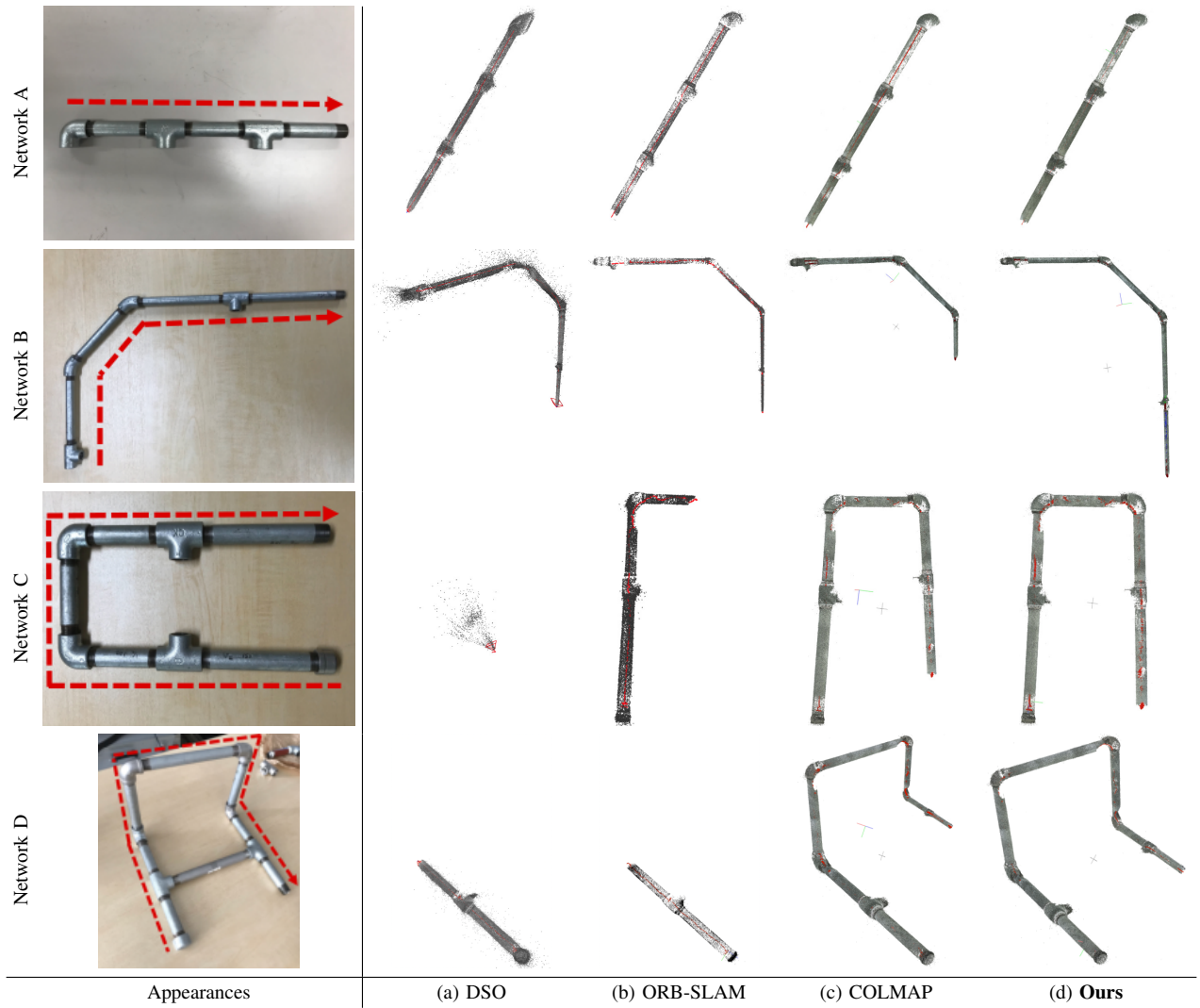
Fig. 6. **Qualitative comparisons.** Each row shows a visual comparison of the 3D models for each pipe network obtained via four relevant methods. Gray dots show the reconstructed scene points, whereas red dots show the estimated cameras.

[32] H. Zhou, B. Ummenhofer, and T. Brox, "DeepTAM: Deep tracking and mapping," in *Proc. ECCV*, 2018.

[33] T. Kishi, M. Ikeuchi, and T. Nakamura, "Development of a peristaltic crawling inspection robot for 1-inch gas pipes with continuous elbows," in *Proc. IEEE/RSJ Conf. on Intelligent Robots and Systems*, 2013.

[34] Y. Gong, R. S. Johnston, C. D. Melville, and E. J. Seibel, "Axial-stereo 3-d optical metrology for inner profile of pipes using a scanning laser endoscope," *International journal of optomechatronics*, vol. 9, no. 3, pp. 238–247, 2015.

[35] P. Hansen, H. Alismail, B. Browning, and P. Rander, "Stereo visual odometry for pipe mapping," in *Proc. IEEE/RSJ Conf. on Intelligent Robots and Systems*, 2011.

[36] S. Esquivel, R. Koch, and H. Rehse, "Reconstruction of sewer shaft profiles from fisheye-lens camera images," in *Joint Pattern Recognition Symposium*, 2009.

[37] Y. Zhang, R. Hartley, J. Mashford, L. Wang, and S. Burn, "Pipeline reconstruction from fisheye images," *Journal of WSCG*, vol. 19, pp. 49–57, 2011.

[38] J. Kunzel, T. Werner, P. Eisert, and J. Waschnewski, "Automatic analysis of sewer pipes based on unrolled monocular fisheye images," in *Proc. WACV*, 2018.

[39] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[40] C. Beder and R. Steffen, "Determining an initial image pair for fixing the scale of a 3d reconstruction from an image sequence," in *Joint Pattern Recognition Symposium*, 2006.

[41] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *PAMI*, vol. 25, no. 8, pp. 930–943, 2003.

[42] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment - A modern synthesis," *Vision Algorithms: Theory and Practice*, pp. 298–372, 1999.

[43] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *PAMI*, vol. 28, no. 8, pp. 1335–1340, 2006.

[44] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[45] D. Lopez-Escogido and L. G. de la Fraga, "Automatic extraction of geometric models from 3D point cloud datasets," in *Proc. CCE*, 2014.

[46] K. Lebeda, J. Matas, and O. Chum, "Fixing the locally optimized RANSAC–full experimental evaluation," in *Proc. BMVC*, 2012.

[47] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *PAMI*, vol. 32, no. 5, pp. 815–830, 2009.

[48] A. R. Widya, A. Torii, and M. Okutomi, "Structure-from-Motion using dense CNN features with keypoint relocalization," *IPSJ Transactions on Computer Vision and Applications*, vol. 10, no. 1, p. 6, 2018.