Visual-Inertial-Laser-Lidar (VILL) SLAM: Real-time Dense RGB-D Mapping for Pipe Environments

Tina Tian[1], Luyuan Wang[1], Xinzhi Yan[1], Fujun Ruan[1], G. Jaya Aadityaa [1], Howie Choset[1], Lu Li[1,*]

***Abstract*— Robotic solutions for pipeline inspection promise enhancement of human labor by automating data acquisition for pipe condition assessments, which are vital for the early detection of pipe anomalies and the prevention of hazardous leakages and explosions. Through simultaneous localization and mapping (SLAM), colorized 3D reconstructions of the pipe's inner surface can be generated, providing a more comprehensive digital record of the pipes compared to conventional vision-only inspection. Designed for generic environments, most SLAM methods suffer limited accuracy and substantial accumulative drift in confined and featureless spaces such as pipelines, due to a lack of suitable sensor hardware and state estimation techniques. In this research, we present VILL-SLAM: a dense RGB-D SLAM algorithm that combines a monocular camera (V), an inertial sensor (I), a ring-shaped laser profiler (L), and a Lidar (L) into a compact sensor package optimized for in-pipe operations. By fusing complementary visual and depth information from the color camera, laser profiling, and Lidar measurement, our method overcomes the challenges of metric scale mapping in conventional SLAM methods, despite its monocular configuration. To further improve localization accuracy, we utilize the pipe geometry to formulate two unique optimization factors that effectively constrain odometer drift. To validate our method, we conducted real-world experiments in physical pipes, comparing the performance of our approach against other state-of-the-art algorithms. The proposed SLAM framework achieved 6.6 times drift improvement with 0.84% mean odometry drift over 22 meters and a mean pointwise 3D scanning error of 0.88mm in 12-inch diameter pipes. This research represents a significant advancement in miniature in-pipe inspection, localization, and mapping sensing techniques. It has the potential to become a core enabling technology for the next generation of highly capable in-pipe robots, capable of reconstructing photo-realistic 3D pipe scans and providing disruptive pipe locating and georeferencing capabilities.**

I. INTRODUCTION

Pipelines, crucial infrastructures supporting human civilization, may experience degradation from various factors like corrosion, geological subsidence, and improper plumbing or digging, leading to economic losses and hazardous incidents. The inspection and maintenance of pipelines are of paramount importance. For inspecting the inner surface of the pipes, one of the most well-adopted techniques is to use closed-circuit television (CCTV) cameras for basic visual inspection, and oftentimes it relies on human operators for time-consuming data collection and video analysis [1]. Due to the monomodal nature of 2D images, such methods may fall short of the objectives to detect and localize anomalies.

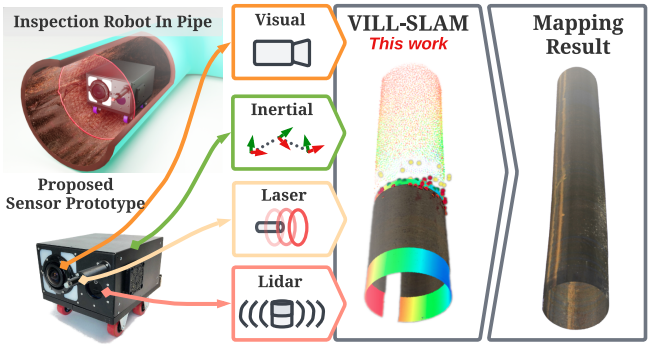

Fig. 1. The in-pipe mapping sensor hardware prototype using the proposed VILL-SLAM method to fuse multiple sensory information, which produces a photo-realistic dense RGB-D reconstruction of a 12-inch pipe segment.

To acquire a more comprehensive digital record of the pipe interior, multi-sensor fusion SLAM techniques from the robotics community can be used to automatically collect sensory data and perform RGB-D reconstructions in pipes, providing a combination of visual, 3D, and georeferencing information for more sophisticated pipe condition monitoring and assessment. Compared to single-sensor odometry and mapping methods such as [2]–[4], fusing data from multiple sensor modalities increases the reliability of state estimation and helps reduce ambiguity during the estimation. Popular methods like [5] [6] make use of an inertial measurement unit (IMU) alongside visual or Lidar odometry. [7], [8] incorporate visual, inertial, and Lidar information to further improve the SLAM performance when there is perceptual degradation or fast motion. These algorithms and the corresponding sensor suites are designed for generic indoor and outdoor environments.

In confined environments such as pipes, however, the sensing hardware and software employed in most conventional methods suffer low localization and mapping accuracy. From a hardware perspective, some methods are entirely unable to operate inside small pipes due to the limitation of minimal sensing range and bulky sensors. Algorithmically, the slower sensor motion in confined spaces can lead to insufficient IMU excitation and thus an incorrect metric scale or IMU bias initial estimation. The lack of visual and geometric features also poses a significant challenge to state estimation. Pipe environments lack distinct 3D geometric features such as edges, planes, and corners compared to daily environments, rendering Lidar odometry methods that heavily rely on these

This work was supported by the Department of Energy Advanced Research Projects Agency-Energy (ARPA-E) under the Rapid Encapsulation of Pipelines Avoiding Intensive Replacement (REPAIR) program.

[1]*The authors are with the Biorobotics Lab, the Robotics Institue at Carnegie Mellon University, Pittsburgh, PA 15213, USA*

[*]Lu Li is the corresponding author. Email: lilu12@andrew.cmu.edu

features inappropriate. Likewise, the scarcity of 2D visual features in pipes may cause feature tracking failure and thus lead to inconsistent scale estimation during visual-based SLAM processes. Methods like [9] and [10] that address this issue by assuming a fixed measured pipe radius to constrain the depth of the observed visual features have nonetheless made an assumption that limits their use to pipes with one specific diameter only. Alternatively, [11] leverages laser profiling to actively determine the local depth, and has shown promising results in improving the accuracy of metric scale estimation and reducing the localization drift.

Leveraging multi-sensor fusion and laser profiling, we have previously proposed a method, VLI-SLAM [12], which integrates single-line structured light [13] and visual-inertial-based approaches [5], and demonstrated the potential of applying it in short-range and confined space SLAM applications. Although tests indicate that this method can create a highly detailed pipe scan while maintaining relatively consistent localization tracking in short distances, we do observe a non-negligible drift over longer pipe segments.

In environments where global positioning information is unavailable, any SLAM algorithm that only relies on spatially or temporally local measurements is likely to experience amplified odometry drift over the long run due to the accumulation of uncorrected dead-reckoning errors. To mitigate this issue, researchers have looked into methods that use the pipe's cylindrical shapes to correct the odometry drift. For example, [14] incorporates a cylindrical constraint into ORB-SLAM2 [2] for long-term drift reduction. However, this work assumes that the pipe is perfectly straight, which is an oversimplified assumption. In this research, we aim to further explore different types of constraints derived from pipe geometry that adapt to the pipe's structural characteristics, and construct algorithms to enhance our previous VLI-SLAM method and optimize it for pipe environments.

Our contributions are summarized as follows:

- A sensor suite design and software framework capable of mapping in compact pipes with real-time localization and photorealistic dense RGB-D mapping capability.
- A sliding-window-based SLAM pipeline with a novel combination of Lidar-based constraints derived from pipe geometric structure for long-term drift reduction.

II. SYSTEM OVERVIEW

A. Hardware System Overview

Our proposed sensor suite design (Fig. 1b) consists of an RGB CMOS camera with a fisheye lens, a MEMS-based 6-axis IMU, a laser projector, and a Realsense L515 Lidar. The laser projector projects a laser plane orthogonal to the camera's optical axis by emitting a thin laser beam towards a conic mirror, and the laser plane forms a ring on the pipe's inner surface. We utilize the alternating-shutter technique described in [12] to strobe the red laser stripe and the illumination LED in synchronization with the image shutter trigger (Fig. 2). This way, we effectively capture the visual frames $\mathcal{I}_v$ containing visual details and the profiling frames $\mathcal{I}_p$ containing a bright laser ring with minimal time gaps. The streaming rate of the visual-laser frame pairs is 30Hz. We perform camera intrinsic calibration with [15], camera-laser calibration with [13], camera-IMU calibration with [16], and camera-Lidar calibration using MATLAB's Lidar Toolbox. The proposed sensor package is compact and attachable to existing actuated in-pipe crawler robots.

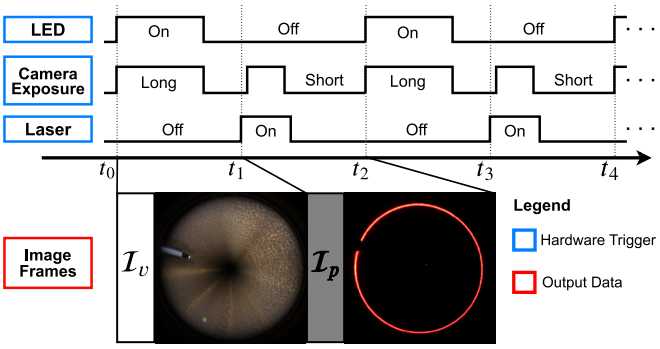

Fig. 2. We capture visual and profiling image frames using one camera by quickly switching between visual and profiling frames using a microcontroller, measuring RGB and depth information in a near-concurrent fashion.

B. Software System Overview

The software architecture diagram is illustrated in Fig. 3. After acquiring data from each physical sensor, a series of front-end preprocessing processes formulate data frames by tracking the visual features, preintegrating IMU measurements [17], triangulating the laser profiler data, and fitting cylinder primitives using the Lidar point cloud. Then the software associates the 2D visual features with depths from the triangulated 3D laser points to bootstrap the SLAM process. With the preprocessed sensory data from the front-end, a sliding-window-based nonlinear optimization process is activated to estimate robot odometry and visual feature depths, where the constraints in the cost function are structured with a factor graph. Using the odometry determined from the state estimator, the 3D laser scans are registered into a colored point cloud map. We also feed the odometry back to register a local Lidar map, which provides global depth information during the visual-depth association.

III. VISUAL-INERTIAL-LASER-LIDAR SLAM

At the core of this method, is a sliding-window-based nonlinear optimization that performs state estimation using visual, inertial, and laser-and-lidar-induced depth measurements. It also incorporates both the pipe's cylindrical structure and available geometric features to reduce odometry drift. Finally, the output is a dense RGB-D scan of the pipe interior and the sensor package's 6-DoF odometry.

A. Sensor Data Preprocessing

The initial stage comprises the pre-processing of raw data obtained from the various sensors. Each data stream is processed independently, and is outlined as follows:

1) Laser detection and triangulation. Given the calibrated laser plane parameters with respect to the camera

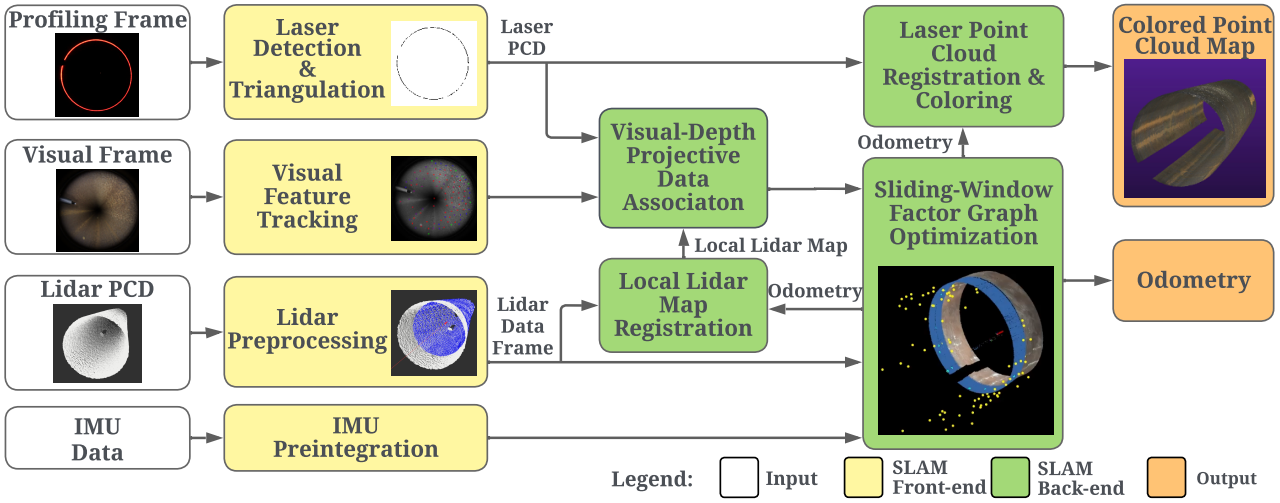

Fig. 3. VILL-SLAM software architecture and information flow chart.

frame, we detect and triangulate the laser points through HSV thresholding in the profiling frames to find the corresponding 3D laser point for each 2D laser pixel following the method described in [12]. In practice, the laser can be robustly extracted even when ambient light is captured in the profiling frames, for example, near pipe entrances and exits.

2) Visual feature tracking through KLT optical flow [5].

3) IMU preintegration, implemented according to [5].

4) Lidar preprocessing Lidar point clouds streaming at 10Hz are cropped to a region of interest between 0.4m and 6m from the sensor, spatially downsampled by a factor of 20, and transformed to the body frame, where the body frame is aligned to the IMU frame. If the remaining points are sufficient, we then attempt to fit a cylinder to the pipe point cloud using RANSAC-based cylinder fitting [18], regardless of whether the captured point cloud is actually cylindrical or not. The output of Lidar preprocessing is a Lidar data frame, which contains the preprocessed Lidar point cloud $\mathcal{L}$, the set of points-on-cylinder $\mathcal{Y} \subseteq \mathcal{L}$, and the corresponding cylinder parameters if the cylinder is found, including the normalized axis vector $\boldsymbol{u}$, an axis point $\boldsymbol{p}$, and the radius r.

B. Visual-Depth Association

After the preprocessing step, the visual features are associated with the corresponding depth data from laser profiling and Lidar. Denote the set of visual features as $\mathcal{F}$. A feature is defined to be a feature-on-laser if any of its observations is close to the laser pixels in adjacent $\mathcal{I}_p$. Denote the set of features-on-laser as $\mathcal{F}_l$. Similarly, a feature $f \notin \mathcal{F}_l$ is a feature-on-Lidar if any of its observations is close to a point in the local Lidar map projected to $\mathcal{I}_v$, and the set of features-on-Lidar is denoted as $\mathcal{F}_L$. For a feature $f_i \in \mathcal{F}_l \cup \mathcal{F}_L$, the observation frame in which the feature pixel is the closest to the laser pixel or the projected Lidar pixel is defined as its primary observation frame c_i^*. If $f_i \notin \mathcal{F}_l \cup \mathcal{F}_L$, its c_i^* is the first observation frame. Note that if a feature $f_i \in \mathcal{F}_l \cup \mathcal{F}_L$, there exists a depth association between the 2D feature and 3D data. Denote its associated depth at c_i^* as $\bar{d}_i$.

C. Estimator Initialization

To bootstrap the SLAM process, a vision-only structure from motion (SfM) is performed to obtain the up-to-scale camera poses and feature positions, followed by visual-inertial alignment, which aligns metric IMU pre- integration with the SfM results [5]. This way, we can obtain a rough gyroscope bias estimation $\hat{\mathbf{b}}_g$ and a scale estimation $\hat{s}$. The gyroscope bias is prone to error due to insufficient rotational excitation, and the scale estimation can be a few orders of magnitude larger than the true scale in an in-pipe scenario. A fine-tuned scale estimation, $\bar{s}$, is computed by multiplying the rough scale by a factor determined by the associated depth $\bar{d}_i$ of each visual feature $f_i \in \mathcal{F}_l \cup \mathcal{F}_L$ (1). $|\cdot|$ denotes the cardinality of a set, $\hat{d}_i$ is the estimated depth of each feature from triangulation. α balances the contribution of laser and Lidar depths to the scale correction and it is set to $|\mathcal{F}_L|/(|\mathcal{F}_l| + |\mathcal{F}_L|)$ if $|\mathcal{F}_L| > 0$ or 1 otherwise.

$$\bar{s} = \alpha \frac{1}{|\mathcal{F}_l|} \sum_{f_i \in \mathcal{F}_l} \frac{\bar{d}_i}{\hat{d}_i} \hat{s} + (1-\alpha) \frac{1}{|\mathcal{F}_L|} \sum_{f_i \in \mathcal{F}_L} \frac{\bar{d}_i}{\hat{d}_i} \hat{s} \tag{1}$$

Using the new scale, we correct each keyframe's position, velocity, and feature positions estimated during visual-inertial alignment and conclude the initialization process. Note that we do not fine-tune the gyroscope bias estimation here since it will be further corrected online during the sliding-window-based optimization.

D. Sliding-Window-Based Factor Graph Optimization

After estimator initialization, we proceed with a sliding window-based tightly-coupled monocular VIO. The full state vector, consisting of robot states $\mathbf{x}$ and landmark states λ in the sliding window, is defined in (2):

$$\begin{aligned} \mathcal{X} &= [\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n, \lambda_0, \lambda_1, ..., \lambda_m] \\ \mathbf{x}_k &= [\mathbf{T}_{b_k}^w, \mathbf{v}_{b_k}^w, \mathbf{b}_a, \mathbf{b}_g], \end{aligned} \tag{2}$$

where n, m are the total number of keyframes and visual features in the sliding window, respectively. λ_i is the inverse

feature depth of f_i in its primary observation frame c_i^*. $\mathbf{b}_a$ and $\mathbf{b}_g$ are the accelerometer and gyro biases [5], $\mathbf{T}_{b_k}^w$ is the pose of the k^{th} body frame with respect to the world frame.

We perform maximum a posteriori estimation of the states $\mathcal{X}$ by minimizing the weighted sum of four factors in a factor graph, alongside the pose prior obtained from the last marginalization [5], as shown in Fig. 4.

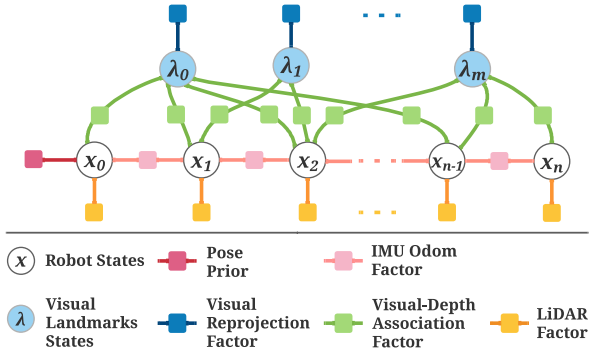

Fig. 4. At the core of our method, a factor graph is used to represent various states with their related constraints within a sliding window.

Visual Reprojection Factor measures the re-projection error of visual features between keyframes. For every feature f_i, we project its estimated 3D coordinates to keyframe c_j in the window and compute the 2D pixel difference with the observation in its primary observation frame c_i^* (3). $\pi_c(\cdot)$ projects a 3D point onto the 2D image and $\pi_c^{-1}(\cdot)$ back-projects a pixel onto the normalized image plane.

$$e_c = \sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{Y}} \left|\left| \pi_c(\mathbf{T}_w^{c_j}\mathbf{T}_{c_i^*}^w \frac{1}{\lambda_i}\pi_c^{-1}(\begin{bmatrix} u_i^{c_i^*} \\ v_i^{c_i^*} \end{bmatrix})) - \begin{bmatrix} u_i^{c_j} \\ v_i^{c_j} \end{bmatrix} \right|\right|^2 \quad (3)$$

Visual-Depth Association Factor leverages the depth association between visual features and laser or Lidar information to further constrain feature depth estimation. For a feature $f_i \in \mathcal{F}_l \cup \mathcal{F}_L$, we compute the residual between the associated depth $\bar{d}_i$ from projective data association and the estimated feature depth, according to (4):

$$e_d = \sum_{f_i\in\mathcal{F}} ||\frac{1}{\lambda_i} - \bar{d}_i||^2 \quad (4)$$

IMU Odometry Factor We follow the IMU residual definition in [5] to estimate $\mathbf{v}_{b_k}^w$, $\mathbf{b}_a$, $\mathbf{b}_g$, and $\mathbf{T}_{b_k}^w$. The IMU factor is used to assist state estimation when low-feature regions are encountered [10].

Lidar Factor is used to constrain the odometry estimation by aligning Lidar frames through Lidar point cloud matching. This alignment is performed between each Lidar frame $\mathcal{L}_k$ in the current sliding window and a reference Lidar frame $\mathcal{L}_{ref}$ captured prior to the window and updated periodically with a fixed update period $\mathcal{T}_{ref}$, and the pose $\mathbf{T}_{b_{ref}}$ of the body frame b_{ref} in which $\mathcal{L}_{ref}$ was captured is known.

We define two candidate Lidar factors: 1) the Lidar cylinder factor, inspired by [14], for long and straight pipe segments and 2) the Lidar iterative-closest-point (ICP) factor when the sensor package approaching geometrically-diverse environments such as pipe with dents or protrusions, fittings, branches, or deformed areas. Depending on the environmental characteristics, the algorithm selects one Lidar factor from these two candidates. The criterion for the selection is termed as the cylindrical structure regularity $R_\mathcal{Y}$, as defined in (5). The larger $R_\mathcal{Y}$ is, the more Lidar points are points-on-cylinder and the environment is more cylindrical.

$$R_\mathcal{Y} = \begin{cases} |\mathcal{Y}|/|\mathcal{L}|, & \text{if } |\mathcal{L}| > 0 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

1) Lidar Cylinder Factor is selected if $R_\mathcal{Y}$ is greater than a threshold $H_\mathcal{Y}$ for both $\mathcal{L}_k$ and $\mathcal{L}_{ref}$, which means that the environment is almost perfectly cylindrical. The key idea is that for two sets of points-on-cylinder, if they capture the same physical cylindrical environment, all points are equidistant from one single cylinder axis. This cylindrical constraint is mathematically formulated in (6), where $\mathcal{Y}_k \in \mathcal{L}_k$ and $\mathcal{Y}_{ref} \in \mathcal{L}_{ref}$. These points-on-cylinder are assumed to lie on the same physical cylinder given $\mathcal{T}_{ref}$, which is empirically set based on the maximum curvature along the pipe and the robot's speed. We empirically set it to 10s. $\boldsymbol{q}_i$ is any point in $\mathcal{Y}_k$. $\boldsymbol{u}^{b_k}$, $\boldsymbol{p}^{b_k}$ and $\boldsymbol{u}^{b_{ref}}$, $\boldsymbol{p}^{b_{ref}}$ are the cylinder axis vectors and axis points in frame b_k and b_{ref}, respectively (Fig. 5). Although the two degrees of freedom about the axial direction of the pipe are underconstrained with this cylinder factor alone, the pose is fully constrained when jointly optimized with other factors.

$$e_\mathcal{Y} = \sum_{k=0}^{n}\sum_{\boldsymbol{q}_i\in\mathcal{Y}_k} \left|\left| \mathbf{T}_{b_{ref}}^w \boldsymbol{u}^{b_{ref}} \times \left(\mathbf{T}_{b_k}^w \boldsymbol{q}_i^{b_k} - \mathbf{T}_{b_{ref}}^w \boldsymbol{p}^{b_{ref}}\right) \right|\right| - ||\mathbf{T}_{b_{ref}}^w \boldsymbol{u}^{b_{ref}}|| \cdot r \quad (6)$$

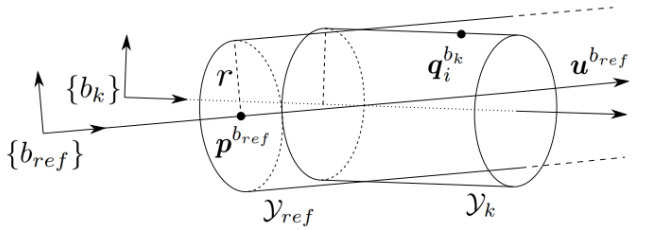

Fig. 5. Lidar cylinder factor.

2) Lidar ICP Factor is selected if $R_\mathcal{Y} \in (0, H_\mathcal{Y}]$ for both $\mathcal{L}_k$ and $\mathcal{L}_{ref}$. First, point-to-point ICP [20] is performed on $\mathcal{L}_k$ and $\mathcal{L}_{ref}$ to find the matching point pairs in the Lidar point clouds. Then, we construct the residual shown in (7) by transforming the corresponding points into the world frame and computing the distance between each pair of correspondences. The ICP can be completed without ambiguity because the threshold check on the cylindrical structure regularity ensures that there exist sufficient geometric features to perform the alignment.

$$e_{icp} = \sum_{\boldsymbol{q}_i\in\mathcal{L}_k\cap\mathcal{L}_{ref}} \left|\left| \mathbf{T}_{b_k}^w \boldsymbol{q}_i^{b_k} - \mathbf{T}_{b_{ref}}^w \boldsymbol{q}_i^{b_{ref}} \right|\right|^2 \quad (7)$$

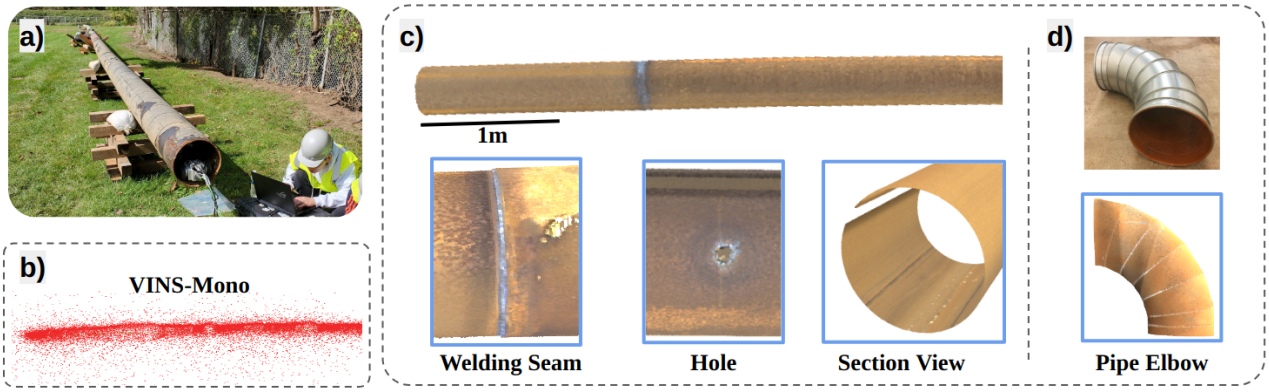

Fig. 6. Test environment and mapped point cloud visualization. (a) Test site with a total pipe length of 50m. (b) Mapping results from VINS-Mono [5]. Note that other tested methods including LOAM [3], FAST-LIO2 [6], ORB-SLAM2 [2] and dense mapping pipelines like SSL-SLAM2 [4] and RGBDTAM [19] all failed to produce reasonable results under same testing conditions. (c) Dense colorized 3D map created by VILL-SLAM (ours) in the pipe test site and the zoom-in views of some regions of interest. (d) 3D map created by VILL-SLAM in a pipe elbow sample.

E. Map Registration

With the odometry from the optimization, we register two maps in the world frame: 1) a dense and colored map of the laser scans and 2) a sparse map of Lidar point clouds. The laser map is used as the final output of the SLAM and is generated in a similar fashion as described in [12], where the color of each laser scan is estimated using adjacent visual frames. The Lidar map is an intermediate product needed by the visual-depth association step. To ensure bounded memory usage, only the Lidar frames captured in the last τ_{ref} are kept in the Lidar map, and the points are downsampled using a voxel grid filter [18]. The Lidar map allows the visual-depth association between a current visual frame and historical Lidar data, which is desirable in long narrow pipes.

IV. EXPERIMENTS

Experiments are conducted to functionally validate our method's dense RGB-D mapping capability and verify that it is able to perform low-drift localization and accurate 3D reconstruction. First, we visually compare the mapping quality of our method against the state-of-the-art SLAM methods. Next, we conduct two experiments to quantitatively analyze the localization and mapping performance. All our experiments are performed in 12-inch diameter metal pipes. No additional lighting besides the LED on the sensor prototype and no fiducials are placed in the pipe.

A. Qualitative Analysis

We visually compared the mapping result of our VILL-SLAM to various state-of-the-art single or multi-sensor SLAM algorithms, including FAST-LIO2 [6], ORB-SLAM2 [2], LOAM [3], VINS-Mono [5] and dense mapping methods like SSL-SLAM2 [4] and RGBDTAM [19]. Among the evaluated methods, all except our methods and VINS-Mono fail to localize and generate a meaningful map of the pipe. Although VINS-Mono is able to generate a map of the pipe, the map is sparse, noisy, and inaccurate due to the incorrect scale estimation in pipes and the lack of dense mapping functionality. In contrast, our VILL-SLAM is able to generate denser and smoother maps (Fig. 6c) thanks to the accurate metric scale estimation through visual-depth association with laser and Lidar. Our output map is also colorized using the alternating-shutter laser profiling method (Sec. III B). In Fig. 6d, we also include a mapping result of a 90° metal duct, whose inner surface is painted to prevent reflection. The photo-realistic point cloud maps of the pipe interior built by VILL-SLAM contain both geometric and visual details of the pipe, which is valuable for downstream surface defect or pipe geometry analysis.

B. Localization Accuracy Evaluation

We collect the ground truth trajectory with a Leica total station (Leica Geosystems AG, Heerbrugg, Switzerland) to track a prism mounted on the top of the sensor package. Since the estimated and the true trajectories are not in the same coordinate frame, we first find the transformation between the world frame set by the state estimator and the total station's frame by aligning the first 10m of the trajectories using a closed-form method described in [21]. Subsequently, the two trajectories are transformed into the same coordinate frame, and the distance between every two matched points is calculated. We compute the absolute trajectory error (ATE) over the trajectory length, which is clipped to 22m.

We compare the localization accuracy of VILL-SLAM against VINS-Mono and VLI-SLAM. Table I shows the performance statistics across 6 trials, where drift is defined as the maximum error over trajectory length. The mean odometry drift of VILL-SLAM is 0.84% and the mean RMSE is 7.62cm, which is 6.6 times drift improvement and 2.8 times RMSE improvement compared to VLI-SLAM. VILL-SLAM also shows a smaller error variance. From this experiment, we verify that our proposed method is able to achieve low drift even over long distances in pipes. On the testing PC with AMD Ryzen 5900x CPU, the average mapping frame rate is 27fps, which is sufficient for real-time mapping. An example of the ATE of VLI-SLAM and VILL-SLAM is shown in Fig. 7. The plot for VINS-Mono is not included because the error is too large compared to the rest.

TABLE I

MEAN AND STANDARD DEVIATION OF LOCALIZATION ERRORS

Metrics	VILL-SLAM (our current)	VLI-SLAM (our prior)	VINS-Mono
RMSE (cm)	**7.62 ± 4.38**	21.67 ± 10.07	1285.71 ± 251.94
Max (cm)	**18.46 ± 7.73**	121.78 ± 40.08	4205.16 ± 452.07
Drift (%)	**0.84 ± 0.35**	5.53 ± 1.82	189.34 ± 20.55

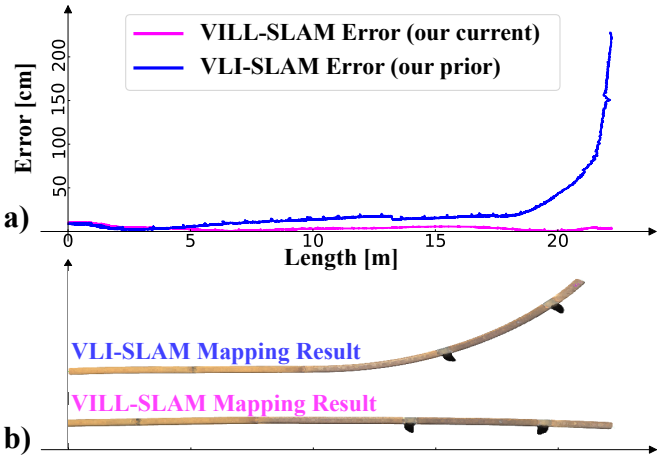

Fig. 7. Example ATE plots of VLI-SLAM and VILL-SLAM. (a) Visualization of the ATE of our current work VILL-SLAM (with Lidar) and our prior method VLI-SLAM (without Lidar) using data collected in one trial. (b) The corresponding 3D maps of the pipe using the two methods.

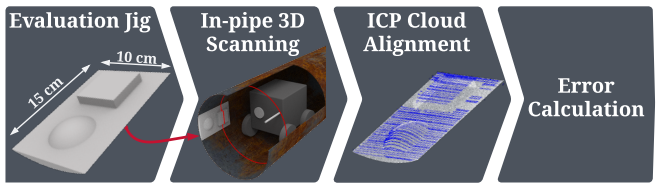

Fig. 8. Workflow of 3D reconstruction accuracy assessment using an evaluation jig and in-pipe 3D scanning experiments.

C. 3D Reconstruction Evaluation.

As shown in Fig. 8, we first use a high-end 3D printer to print custom evaluation jigs, whose ground-truth point clouds are exported from the CAD software. Since the 3D printer has relatively high printing accuracy, the printing error is considered negligible. After installing the evaluation jig onto the pipe's inner surface, we scan the pipe and the jig using the sensor prototype. Finally, the point cloud map from SLAM and the ground-truth point cloud are aligned using ICP, and the point-to-point L2 distance error is computed. Across 4 trials using 2 jig designs, the average error is 0.88mm per point, indicating that our scanner can produce maps with sub-millimeter grade local scanning accuracy.

V. CONCLUSION AND DISCUSSION

In this paper, a confined-space mapping algorithm for accurate localization and 3D reconstruction (mapping) inside narrow pipelines is introduced, which tightly couples a monocular camera, an IMU, a laser profiler, and a Lidar as the entirety of the basic sensor suite. This framework includes the fusion of the two redundant yet complementary depth information gathered from the laser profiler and the Lidar, which prevents accumulative drift during long-distance travel inside a featureless and GPS-denied pipe environment. To further improve the localization accuracy, we also take advantage of the cylindrical pipe structure and formulate two unique optimization factors, Lidar Cylinder Factor for long straight pipe segments and Lidar ICP Factor for geometrically-diverse environments, further constraining the state estimation process. Lastly, the real-world experimental results prove the proposed VILL-SLAM outperformed conventional state-of-the-art SLAM methods in terms of both localization accuracy and mapping accuracy during in-pipe robotic inspection missions.

Experiments indicate that our method is capable of producing sub-millimeter grade photo-realistic 3D reconstruction in a real-time fashion, with an average of 0.84% localization drift and 0.88mm per-point local scanning error, in a 12-inch diameter pipe over 22-meter mapping distance.

While our method offers many advantages, there are several limitations. First, our work has not been fully tested in pipes with a combination of multiple straight and bendy segments because our existing in-pipe crawler robot is incapable of traversing in such complex pipelines. We plan to perform rigorous testing with an upgraded robot in more diverse pipeline configurations and further verify the robustness of our algorithm. Additionally, we observe that the localization accuracy of our method is sensitive to calibration quality, especially from camera-laser extrinsic calibration, in which even small perturbations will result in major mapping errors. One potential solution for this problem is to implement an online calibration procedure other than the current one-off pre-calibration step. Thirdly, the generalizability of the proposed method can be improved by devising other constraints derived from geometric structures in other pipe environments with different cross-section shapes, such as rectangular ducts. Last but not least, it is also interesting to enable loop closure for pipe networks with loops.

For the next phase of this work, we are actively developing an articulated and modular compact pipe crawler robot with higher mobility in narrow pipe environments. Integrating it with the proposed sensor package and SLAM software, we aim to deploy them in various pipe types and damage conditions with our collaborators in the energy and infrastructure industries and evaluate their performance and potential market value proposition.

ACKNOWLEDGMENT

The authors would like to express sincere gratitude to Peoples Natural Gas Co. for providing the testing facilities.

REFERENCES

[1] Z. Liu and Y. Kleiner, "State of the art review of inspection technologies for condition assessment of water pipes," *Measurement: Journal of the International Measurement Confederation*, vol. 46, pp. 1–15, 2013.

[2] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[3] J. Zhang and S. Sanjiv, "Loam: Lidar odometry and mapping in real-time." *Robotics: Science and Systems*, vol. 2, nov 2014.

[4] H. Wang, C. Wang, and L. Xie, "Lightweight 3-d localization and mapping for solid-state lidar," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1801–1807, 2021.

[5] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[6] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-lio2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2053–2073, 2022.

[7] S. Zhao, H. Zhang, P. Wang, L. Nogueira, and S. Scherer, "Super odometry: Imu-centric lidar-visual-inertial estimator for challenging environments," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 8729–8736.

[8] J. Lin and F. Zhang, "R3live: A robust, real-time, rgb-colored, lidar-inertial-visual tightly-coupled state estimation and mapping package," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 10 672–10 678.

[9] P. Hansen, H. Alismail, P. Rander, and B. Browning, "Monocular visual odometry for robot localization in lng pipes," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3111–3116, 2011.

[10] R. Summan, W. Jackson, G. Dobie, C. Macleod, C. Mineo, G. West, D. Offin, G. Bolton, S. Marshall, and A. Lille, "A novel visual pipework inspection system," *AIP Conference Proceedings*, vol. 1949, 2018.

[11] K. Matsui, A. Yamashita, and T. Kaneko, "3-d shape measurement of pipe by range finder constructed with omni-directional laser and omni-directional camera," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2537–2542, 2010.

[12] D. C. et al., "Visual-laser-inertial slam using a compact 3d scanner for confined space," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5699–5705, 2021.

[13] D. Cheng, H. Shi, M. Schwerin, M. Crivella, L. Li, and H. Choset, "A compact and infrastructure-free confined space sensor for 3d scanning and slam," in *2020 IEEE SENSORS*, 2020, pp. 1–4.

[14] R. Zhang, M. H. Evans, R. Worley, S. R. Anderson, and L. Mihaylova, "Improving slam in pipe networks by leveraging cylindrical regularity," in *Towards Autonomous Robotic Systems*, C. Fox, J. Gao, A. Ghalamzan Esfahani, M. Saaj, M. Hanheide, and S. Parsons, Eds. Springer International Publishing, 2021, pp. 56–65.

[15] L. Heng, B. Li, and M. Pollefeys, "Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1793–1800.

[16] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1280–1286.

[17] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2017.

[18] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 1–4.

[19] A. Concha and J. Civera, "Rgbdtam: A cost-effective and accurate rgb-d tracking and mapping system," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 6756–6763.

[20] P. Besl and H. McKay, "A method for registration of 3d shapes," *IEEE Trans. Pattern Anal*, pp. 230–256, 1992.

[21] B. K. Horn. (1987) Closed-form solution of absolute orientation using unit quaternions.