2024 IEEE International Conference on Robotics and Automation (ICRA)
May 13-17, 2024. Yokohama, Japan

COIN-LIO: Complementary Intensity-Augmented LiDAR Inertial Odometry

Patrick Pfreundschuh, Helen Oleynikova, Cesar Cadena, Roland Siegwart, and Olov Andersson

2024 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3503-8457-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRA57147.2024.10610938

***Abstract*— We present COIN-LIO, a LiDAR Inertial Odometry pipeline that tightly couples information from LiDAR intensity with geometry-based point cloud registration. The focus of our work is to improve the robustness of LiDAR-inertial odometry in geometrically degenerate scenarios, like tunnels or flat fields. We project LiDAR intensity returns into an image, and present a novel image processing pipeline that produces filtered images with improved brightness consistency within the image as well as across different scenes. We effectively leverage intensity as an additional modality, using our new feature selection scheme that detects uninformative directions in the point cloud registration and explicitly selects patches with complementary image information. Photometric error minimization in the image patches is then fused with inertial measurements and point-to-plane registration in an iterated Extended Kalman Filter. The proposed approach improves accuracy and robustness on a public dataset. We additionally publish a new dataset, that captures five real-world environments in challenging, geometrically degenerate scenes. By using the additional photometric information, our approach shows drastically improved robustness against geometric degeneracy in environments where all compared baseline approaches fail.**

I. INTRODUCTION

Recent advances in 3D Light Detection and Ranging (LiDAR) have decreased both the size and price of these sensors, enabling them to be used by a wider range of robots. At the same time, new LiDAR-based state estimation approaches such as FAST-LIO2 [1] have increased the accuracy and robustness while decreasing the computational cost, making 3D LiDAR one of the most popular choices for mobile robot sensors. However, even these LiDAR-Inertial Odometry (LIO) approaches struggle in geometrically degenerate environments, such as tunnels and flat fields.

In most geometrically uninformative scenes, the texture of the environment still offers some visual information. While several works [2–5] fuse camera information with LiDAR to take advantage of this complementary data, this requires additional sensors, accurate extrinsic calibration, and time synchronization. Cameras will also not work in the absence of ambient light, which limits their applicability. However, in addition to range measurements, LiDARs provide the measured signal strength of each reflected point (intensity). For modern mechanically rotating multi-layer LiDARs, this signal can be projected into a dense image, which allows the LiDAR to operate as an active camera without external illumination. Images and point clouds are time-synchronized and the extrinsics are known, which drastically simplifies their use compared to a combination of LiDAR with cameras.

These intensity images contain texture information about the environment, which can be used for pose estimation.

Authors are with Autonomous Systems Lab, ETH Zurich, e-mail: patripfr@ethz.ch. This work was supported by Swiss National Science Foundation's NCCR DFab P3.

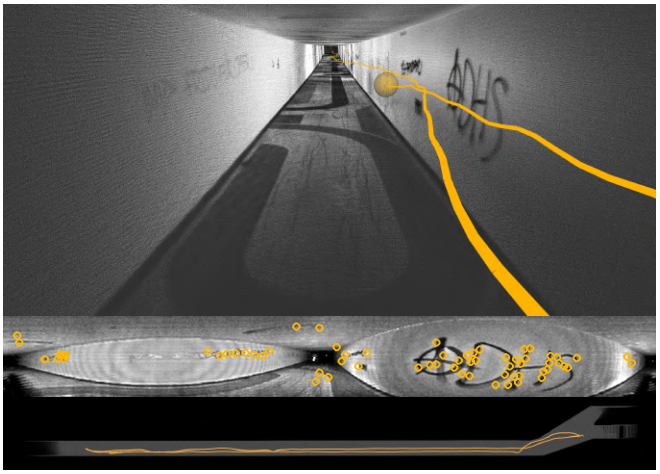

Fig. 1. *Top*: Accumulated point cloud colorized by intensity and trajectory (orange) resulting from COIN-LIO. Our approach achieves accurate odometry despite geometric degeneracy along the tunnel, resulting in clearly visible correct ground and wall markings. *Mid*: Filtered intensity with tracked features (orange). *Bottom*: Top view of the resulting point cloud (gray) and trajectory (orange) from the tunnel.

Compared to camera images, intensity images suffer from poor Signal-to-Noise Ratio, lower resolution, strong rolling shutter effects, and a different projection model from traditional pinhole cameras, making it difficult to directly apply existing visual odometry methods. While several works have used intensity to improve pose estimation [6–8], they perform no filtering to improve image quality and do not combine the information from geometry and intensity in a complementary way, which limits performance in geometrically degenerate scenes as shown in our experiments.

To this end, we present COmplementary INtensity-Augmented LIO, a robust, real-time LIO framework, that couples geometric registration with photometric error minimization for increased robustness. We improve upon related work by introducing a filtering method to increase brightness consistency in the intensity image and a feature selection scheme that adds features with complementary information to the degenerate geometry of the scene. This feature selection is important as parts of the scene (like edges of a tunnel) are often also visually uninformative along the geometrically degenerate direction. The complementary information vastly increases the robustness of the combined method in geometrically degenerate scenarios while keeping or improving performance in geometry-rich scenes.

We found a lack of 3D LiDAR datasets focusing on scenarios with degenerate geometry. To this end, we created the ENWIDE dataset, which captures five real-world ENvironments WIth large sections of DEgenerate geometry and recorded ground truth positions from a high-accuracy laser scanner. We hope that by providing this data to the commu-

nity along with our open-sourced code implementation[1], we can fuel further advances in robust LiDAR-inertial odometry.

The main contributions of our work can be stated as follows: *(1)* we present a LiDAR-intensity image processing pipeline as well as a geometrically complementary feature selection scheme that enables detection and tracking of salient features with complementary information to the geometry-based measurements, *(2)* we show that our approach effectively leverages LiDAR intensity to improve robustness and performance of LIO in geometrically degenerate scenes, *(3)* we provide a real-world dataset, ENWIDE[2], that contains ten sequences in five scenes of diverse geometrically degenerate environments, with accurate position ground truth.

We present our contributions in a combined system with geometry-based LIO, based on FAST-LIO2 [1], and show superior performance on a standard dataset and ENWIDE, over geometry-only and geometry-and-intensity-based methods.

II. Related Work

A. LiDAR (Intertial) Odometry

Common LiDAR-based odometry approaches are based on registration of a measured point cloud against a map that is built during operation. For many years, the standard approach for LiDAR Odometry (LO) was LOAM [9] which extracts points on edges and planes for registration. This works well in structure-rich environments, but edge and plane points are often not expressive enough to perform robustly in geometrically challenging scenarios. KISS-ICP [10] avoids feature selection and directly registers a voxel-downsampled point cloud with point-to-point ICP which showed improved performance in unstructured environments. X-ICP [11] explicitly detects degenerate directions in the registration but relies on an auxiliary state estimate. The use of inertial measurements in LIO approaches has shown a large increase in robustness, as it helps to remove ego-motion distortion from the point cloud and provides an initial guess for the registration. LIO-SAM [12] fuses Inertial Measurement Unit (IMU) measurements in a factor-graph [13, 14] with edge and plane feature matching against submaps. FAST-LIO [15] presents an efficient formulation of the Kalman Filter update that enables the alignment of every scan against the continuously built map in real-time. The authors switch from feature matching to raw points with point-to-plane ICP in its successor [1] that achieves state-of-the-art performance.

B. Intensity Assisted Odometry

Several approaches use intensity as a similarity metric and integrate it into a weighted ICP [16, 17] or use high-intensity points as an additional feature class [18–20], but due to their limited map resolution, these approaches cannot capture fine-grained details. Early works [21–24] have shown that LiDAR intensities can create lighting invariant images that can be used for visual odometry. However, as they only match intensity image features, these approaches do not leverage the geometric information efficiently. Similarly, the approaches presented in [8, 25] detect and match image features in the intensity image for registration. However, in geometrically degenerate cases such features are often sparse and most of the geometric information is neglected, resulting in inferior performance. In these approaches, the intensity only influences the point correspondences but does not directly provide a gradient in the optimization. In contrast, in MD-SLAM [6], the photometric error of the intensity image is optimized together with a range and normal image. Unlike our work, they do not use the IMU or perform motion undistortion. They also use the entire dense image instead of sparse informative patches. The approach closest to our work is RI-LIO [7]; similar to us, they integrate photometric error minimization into the iterated Extended Kalman Filter (iEKF) [26] of [1] but use reflectivity instead of intensity. They randomly downsample the point cloud and project single points into the reflectivity image for the photometric components. However, relevant information is typically not distributed homogeneously in images but concentrated in specific salient regions. Instead of single random pixels at a low resolution, we specifically select geometrically complementary, salient high-resolution patches from a filtered image and continuously assess the feature quality. This leads to superior performance in difficult geometrically-deficient scenarios compared to existing approaches.

[1] https://github.com/ethz-asl/coin-lio

[2] https://projects.asl.ethz.ch/datasets/enwide

III. Method

COIN-LIO adopts the tightly-coupled iEKF presented in FAST-LIO2 for point-to-plane registration and extends it with photometric error minimization. Due to space limitations, we do not review FAST-LIO2 [1, 15] and focus on the photometric component. We process intensity images from point clouds using a novel filter that improves brightness consistency and reduced sensor artefacts. We specifically select image features that provide information in uninformative directions of the point cloud geometry. The feature management module examines the validity of tracked features and detects occlusions. Finally, we integrate the photometric residual into the Kalman Filter.

A. Definitions

We define a fixed global frame (G) at the initial pose of the IMU (I). The transformation from LiDAR frame (L) to IMU frame is assumed to be known as $\mathbf{T}_{IL} = (\mathbf{R}_{IL}, {}_I\mathbf{p}_{IL}) \in SE(3)$. We define the robot's state as $\mathbf{x} = [\mathbf{R}_{GI}, {}_G\mathbf{p}_{GI}, {}_G\mathbf{v}_I, \mathbf{b}^a, \mathbf{b}^g, {}_G\mathbf{g}]$, where $\mathbf{R} \in SO(3)$ denotes orientation, $\mathbf{p} \in \mathbb{R}^3$ is the position, $\mathbf{v} \in \mathbb{R}^3$ describes linear velocity, and $\mathbf{b}^a, \mathbf{b}^g \in \mathbb{R}^3$ indicate accelerometer and gyro biases. The LiDAR frame at t_j is denoted as L_j. Each LiDAR scan consists of points recorded during one full revolution $\mathcal{P} = \{{}_{L_j}\mathbf{p}_j, j = 1, ..., k\}$, with $t_j \leq t_k$.

B. IMU Prediction and point cloud undistortion

We adopt the Kalman Filter prediction step according to FAST-LIO2 [1] by propagating the state using IMU measurement integration from t_j to t_k. Similarly, we calculate the ego-motion compensated, undistorted points at the latest timestamp t_k as: ${}_{L_k}\mathbf{p}_j = \mathbf{T}_{L_kI_k}\mathbf{T}_{I_kI_j}\mathbf{T}_{I_jL_j}\mathbf{p}_j$.

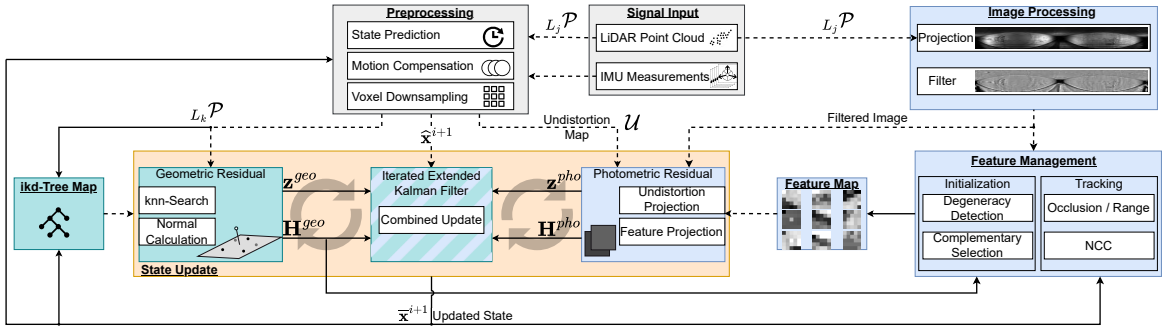

Fig. 2. System Overview: The input point cloud is used geometrically (green) for map registration and as a projected image (blue) for photometric error minimization. Both residuals are combined in an iterated update (orange). We use the registration Jacobian to find uninformative directions in the geometry and select features with complementary image information (right bottom). Lines indicate information flow *before* (- - -) and *after* (—) the update step.

C. Image Projection Model

A point ${}_L\mathbf{p}_j = [x_j, y_j, z_j]$ can be projected to image coordinates using a spherical projection:

$$ {}_C\mathbf{p}_j = \Pi({}_L\mathbf{p}_j) = \begin{bmatrix} f_x\phi + c_x \\ f_y\theta + c_y \end{bmatrix} = \begin{bmatrix} \frac{-w}{2\pi}\operatorname{atan2}(\frac{y_j}{x_j}) + \frac{w}{2} \\ \frac{-h}{\Theta_{fov}}\arcsin(\frac{z_j}{R_j}) + \frac{h}{2} \end{bmatrix} = \begin{bmatrix} u_j \\ v_j \end{bmatrix} \quad (1) $$

with $R = \sqrt{L^2 + z^2}$, $L = \sqrt{x^2 + y^2} - r$, as illustrated in Figure 3. The vertical field of view (FOV) is represented as Θ_{fov}, and w and h denote the horizontal and vertical resolution of the LiDAR. This model assumes a constant elevation angle spacing between subsequent beams. However, for manufacturing reasons, most LiDARs have an irregular spacing which causes empty pixels in the spherical image [7]. While we still use eq. (1) to calculate the Jacobian in eq. (8), we directly use laser beam and encoder value to create the image. We compensate the horizontal offset similar to [7], but use a constant value for all beams. We keep a list of all beam-elevation angles $\Theta_L = \{\theta_1, ..., \theta_h\}$ from the calibration of the LiDAR. When we project a feature point into the image, we calculate θ_f and find the beams above and below in Θ_L to interpolate the subpixel coordinate.

D. Image Processing

The irregular elevation angle spacing between the beams causes horizontal line artefacts in the intensity image. They are less apparent in structure-rich scenes, but dominate the image in environments with little structure. As they occur at a regular row-frequency we design a finite impulse response filter to remove them. First, we use a highpass filter vertically with a cutoff just below the line frequency. Apart from the lines, the highpass signal also contains relevant image content at this frequency. We therefore apply a lowpass filter horizontally to it, which isolates the lines as relevant image signals appear at a higher horizontal frequency. Finally, we subtract the isolated signal from the intensity image. As intensity values depend on the reflectivity of the surface as well as the distance and incidence angle, the intensity is lower in areas that are farther away from the sensor.

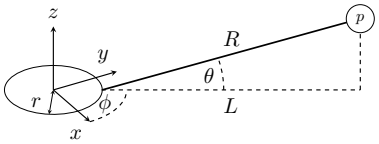

Fig. 3. Projection model. The offset between LiDAR origin and laser emitter is denoted as r. A measured point is depicted on the top right (p).

LiDARs such as the Ouster also report compensated reflectivity signals, which is used in [7], but the influence of the incidence angle remains. We propose a different approach to achieve consistent brightness throughout the image.

The brightness level varies smoothly throughout the image, as average distance and incidence angle are typically driven by the global structure of the scene instead of small geometric details. We thus build a brightness map $I_b(u, v)$ by averaging the intensity values in a large window. To achieve consistent exposure throughout the image, we normalize the pixel values using the brightness map and scale them to values in $[0, 255]$ using a constant factor s_i:

$$ I_F(u, v) = s_i \cdot \frac{I(u, v)}{I_b(u, v) + 1} \quad (2) $$

Finally, we smooth the image using a 3×3 Gaussian kernel to reduce noise. We provide explanatory images in Figure 4.

E. Geometrically Complementary Patch Selection

We select and track 5×5 pixel patches which has shown better convergence compared to single pixels [27]. In contrast to prior works that select features randomly [7] or based on visual feature detectors [8, 25], we follow an approach inspired by [28]. We select candidate pixels with an image gradient magnitude above a threshold and perform a radius-based non-maximum suppression. This approach does not rely on corner features which is favourable for

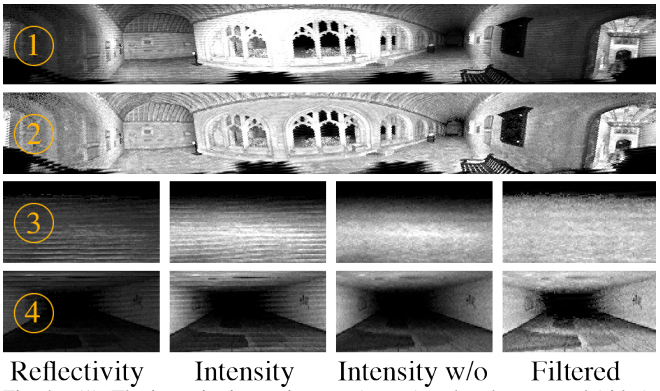

Fig. 4. (1): The intensity image is over- (center) and under-exposed (sides). (2): Our filtered image has consistent brightness across the image. (3) & (4): Detail views from a grass field (3) and tunnel (4). The reflectivity image is under-exposed and does not show the ground markings (4). The intensity suffers from strong line artefacts that dominate the texture (3). Our filter removes the line artefacts (Intensity w/o). Our brightness compensation produces consistent exposure and shows details at larger range (ground markings in (4), grass texture in (3)).

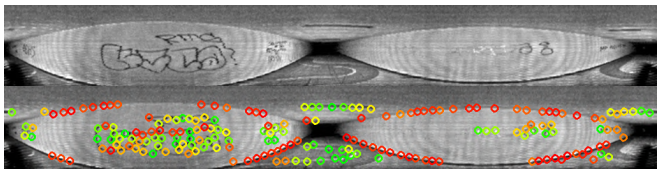

Fig. 5. *Top*: Example frame *Bottom*: Features are colored by contribution strength in the uninformative axis along the tunnel (increases from red to green). Uninformative features along the tunnel edges are correctly marked in red, while features with strong gradients along the tunnel show up green.

low-texture images. Candidate pixels are mostly detected on shape discontinuities in the 3D scene such as edges and corners, or on changes in surface reflectivity, e.g. from ground markings or vegetation. The information from intensity Jacobians of pixels on shape discontinuities often overlaps with the information that is already captured in the point-to-normal registration. We thus aim to select candidates that give additional information to efficiently leverage the multi-modality.

To detect uninformative directions in the point cloud registration, we follow the information analysis presented in X-ICP [11]. We calculate the principal components of the Hessian $\mathbf{H}^{geo^T}\mathbf{H}^{geo}$ of the point-to-plane terms. A direction is then detected as uninformative if the accumulated filtered contribution is below a threshold. We refer the reader to [11] for more details. We analyze the translational components and denote the set of uninformative directions V_t. If all directions are informative, we insert vectors along the coordinate axes to promote equally distributed gradients. We calculate the second image moment M [29] and use its strongest eigenvector $\mathbf{v}_{patch}$ to approximate the patch gradient, which is more stable than pixel gradients. We then calculate how the projected image coordinate changes, if the point is perturbed along a direction using eq. (7):

$$\mathbf{d}_{p_i} = \frac{\partial \Pi(_{L_j}\mathbf{p}_j)}{\partial_{L_j}\mathbf{p}_j} \cdot \mathbf{v}_{t,i} \in \mathbb{R}^2, \forall\, \mathbf{v}_{t,i} \in V_t \tag{3}$$

We select features where shifting the point along an uninformative 3D direction results in a 2D coordinate shift in an informative image direction. We therefore project the projection gradient $\mathbf{d}_{p_i}$ onto the informative direction $\mathbf{v}_{patch}$ of the patch to calculate its directional contribution c_i. As the magnitude of the projection gradient increases with decreasing range, which would favor the selection of points close to the sensor, we use the normalized gradient instead:

$$c_i = \frac{\mathbf{d}_{p_i} \cdot \mathbf{v}_{patch}}{||\mathbf{d}_{p_i}||} \tag{4}$$

For each direction in V_t, we select the patches with the strongest contribution. We visualize the results in Figure 5.

F. Feature Management

We initialize each point in a patch separately at its global position using the current pose estimate. Different from visual odometry approaches [27], where one position is assigned to the whole patch, this allows us to project each point in the patch separately. Using high-resolution patches we can capture fine-grained details in contrast to prior works [7, 17] which only store a single value per voxel-grid cell. To reduce computational load, we limit the number of tracked patches. After each update step, we assess the feature patch validity. To detect occlusions, we compare the predicted and measured range for each point in the patch and discard all points in it if the difference is above a threshold. We also remove patches below a minimum or above a maximum range. Additionally, we calculate the normalized cross-correlation (NCC) between the tracked and measured patch and remove it if the NCC is below a threshold. We only track features over a maximum amount of frames to reduce error accumulation and to encourage the initialization of new features. We avoid overlapping features by enforcing a minimum distance between new and tracked features.

G. Photometric Residual & Kalman Update

We minimize photometric errors between tracked and currently observed points. The error is computed by projecting tracked points into the current image and comparing current intensity values to the patch:

$$z^{pho} = I_c(\Pi(_{L_j}\mathbf{p}_f)) - i_f \tag{5}$$

As rotating LiDARs record individual points sequentially, the pixels inside the intensity image are measured at different times and different poses. For the projection we therefore need to calculate the position of the tracked point in the distorted LiDAR frame L_j:

$$_{L_j}\mathbf{p}_f = \mathbf{T}_{L_jI_j}\mathbf{T}_{I_jI_k}\mathbf{T}_{I_kGG}\mathbf{p}_f \tag{6}$$

However, this is dependent on $\mathbf{T}_{I_jI_k}$, which in turn depends on the unknown time t_j itself. RI-LIO solves this by using a kNN-search in a kD-tree. However, this is only computationally feasible at a low resolution. We thus propose a projection-based solution. Given the undistorted point cloud, we can approximate which volumetric slices of the environment were captured at which timestamp. Therefore, we build an undistortion map by projecting the undistorted point cloud into an image and assign each pixel the index of the corresponding point: $\mathcal{U}(\Pi(_{L_k}\mathbf{p}_j)) = j$.

To find the corresponding index for the feature point, we project it to the undistortion map, which is drastically cheaper than kD-tree-search and thus applicable to the full resolution point cloud. Given the index, we find the respective timestamp and $\mathbf{T}_{I_jI_k}$ to calculate eq. (6) and eq. (5).

The resulting Jacobian $\mathbf{H}^{pho}$ is calculated as:

$$\mathbf{H}^{pho}_j = \frac{\partial \mathcal{I}[_C\mathbf{p}_j]}{\partial_C\mathbf{p}_j} \cdot \frac{\partial \Pi(_{L_j}\mathbf{p}_j)}{\partial_{L_j}\mathbf{p}_j} \cdot \frac{\partial_{L_j}\mathbf{p}_j}{\partial \tilde{\mathbf{x}}} \tag{7}$$

$$\frac{\partial \mathcal{I}[_C\mathbf{p}_j]}{\partial_C\mathbf{p}_j} = \begin{bmatrix} \frac{-f_x y}{x^2+y^2} & \frac{f_x x}{x^2+y^2} & 0 \\ -\frac{f_y xz}{LR^2} & -\frac{f_y yz}{LR^2} & \frac{f_y L}{R^2} \end{bmatrix} \tag{8}$$

$$\frac{\partial_{L_j}\mathbf{p}_j}{\partial \tilde{\mathbf{x}}} = (\mathbf{R}_{L_jL_k}\mathbf{R}_{L_kI}) \begin{bmatrix} [\mathbf{R}_{IG}(_G\mathbf{p}_j - _G\mathbf{p}_{GI})]_\times & -\mathbf{R}_{IG} & \mathbf{0} \end{bmatrix} \tag{9}$$

$\frac{\partial \mathcal{I}[_C\mathbf{p}_j]}{\partial_C\mathbf{p}_j}$ is the image gradient from neighboring pixels. We stack the point-to-plane (^{geo}) and photometric (^{pho}) terms into a combined residual vector ($\mathbf{z}$) and Jacobian ($\mathbf{H}$). The scaling factor σ compensates for the different error magnitudes between geometric and photometric residuals:

$$\mathbf{H} = \left[\mathbf{H}_1^{geo^T}, \cdots, \mathbf{H}_m^{geo^T}, \lambda \cdot \mathbf{H}_1^{pho^T}, \cdots, \lambda \cdot \mathbf{H}_n^{pho^T}\right]^T$$

$$\mathbf{z}_k^\kappa = \left[z_1^{geo}, \cdots, z_m^{geo}, \lambda \cdot z_1^{pho}, \cdots, \lambda \cdot z_n^{pho}\right]^T \quad \mathbf{R} = \text{diag}\left[\sigma\right]$$

We use the formulas provided in [1] to update the state:

$$\mathbf{K} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \tag{10}$$

$$\widehat{\mathbf{x}}_k^{\kappa+1} = \widehat{\mathbf{x}}_k^\kappa \boxplus \left(-\mathbf{K}\mathbf{z}_k^\kappa - (\mathbf{I} - \mathbf{K}\mathbf{H})\left(\mathbf{J}^\kappa\right)^{-1}\left(\widehat{\mathbf{x}}_k^\kappa \boxminus \widehat{\mathbf{x}}_k\right)\right) \tag{11}$$

IV. Experimental Results

We quantitatively compare our proposed pipeline with several state-of-the-art approaches as baselines: KISS-ICP [10], LIO-SAM [12] and FAST-LIO2 [1] represent widely used LO and LIO algorithms. Similar to our approach, MD-SLAM [6], Du and Beltrame [8] and RI-LIO [7] also use intensity or reflectivity information. We use the Newer College Dataset [30] as a public baseline. To evaluate robustness in low-structured environments, we additionally provide and evaluate on a new dataset of geometrically degenerate scenes that is presented in Section IV-B. We calculated the absolute translational error (ATE) and the relative translational error (RTE) over segments of 10 m using the evo library[3]. We declare approaches with an RTE that is larger than 20% as failed (indicated by $\times$) and do not report their ATE, as the required alignment between estimated and ground truth trajectories is not meaningful if the estimated trajectory differs too much from the ground truth. Apart from sensor extrinsics, calibrations, and minimum range (to adapt for narrow scenes), we used the default parameters that were provided by the baseline approaches. We slightly increased the reflectivity covariance parameter in RI-LIO, as the default value caused divergence in all tested sequences.

A. Newer College Results

The Newer College Dataset [30] uses a hand-held 128-beam Ouster OS0. We present the results in Table I. In the *Cloister* sequence, which contains large structures and slow motions, all approaches achieve a low ATE. In *Quad-Hard*, aggressive rotations occur. Due to the absence of accurate ego-motion compensation, the LO approaches perform worst. Our approach achieves the lowest ATE, which confirms that our computationally cheap image motion-compensation method is effective. The *Stairs* sequence causes most approaches to diverge. They use spatial downsampling of the point cloud to achieve real-time performance, which in the case of this narrow stairway removes too much information. While our approach uses the same downsampling for the geometric part, it achieves robust and accurate performance thanks to the photometric component. This unveils an inherent benefit of image-based intensity augmentation: fixed-size patches in the image implicitly capture different amounts of volume depending on the point distance. Thereby, projected images automatically have an adaptive resolution at a constant cost, contrary to the increased cost that would result from the required higher voxel resolution to capture the same information. While RI-LIO uses information from reflectivity images, its random feature selection fails to extract salient information and therefore diverges. In contrast, the dense approach in MD-SLAM does not fail but is outperformed by our approach. We perform slightly better than FAST-LIO2 on the longest and geometry-rich *Park* dataset, showing that the intensity features can also improve performance in non-degenerate scenarios. We also evaluate our runtime on the *Park* sequence. On average, our approach consumes 29.7 ms per frame (33 Hz) on an Intel i7-11800H mobile CPU, of which only 6.2 ms are spent on the photometric components, which shows that the main computational cost results from the conventional geometric approach.

[3] https://github.com/MichaelGrupp/evo

TABLE I
Newer College Dataset
Absolute Trajectory Error (RMSE) (*m*) / Relative Error (%)

Method Length *(m)*	Quad-Hard 234.81	Cloister 428.79	Stairs 57.04	Park 2396.20
KISS-ICP [10]	0.324 / 1.88	0.297 / 2.07	× / 32.48	2.871 / 1.06
MD-SLAM [6]	19.639 / 12.36	0.360 / 2.73	0.340 / 6.21	96.797 / 23.03
Du and Beltrame [8]	18.506 / 16.432	59.544 / 19.274	× / 26.121	× / 42.717
LIO-SAM [12]	0.299 / 2.380	0.145 / 1.032	× / 5122.320	1.566 / 2.064
FAST-LIO2 [1]	0.049 / **0.26**	**0.078** / **0.23**	× / 3497.22	0.310 / 0.59
RI-LIO [7]	0.237 / 1.04	0.285 / 1.33	× / 16877.28	89.289 / 5.00
Ours	**0.046** / 0.29	**0.078** / 0.28	**0.102** / **0.74**	**0.287** / **0.54**

B. ENWIDE Dataset

As geometrically degenerate environments are barely represented in existing open-sourced datasets, we created a new dataset with long segments of real-world geometric degeneracy (Fig. 6). Using a hand-held Ouster OS0 128 beam LiDAR with integrated IMU, we recorded five distinct environments: *Tunnel* (urban, indoor), *Intersection* (urban, outdoor), *Runway* (outdoor, urban), *Field* (outdoor, nature), *Katzensee* (outdoor, nature). All sequences contain long sections of geometric degeneracy but start and end in well-constrained areas. *Tunnel/Intersection/Runway* sequences contain strong intensity features, *Katzensee/Field* contain few salient features. For each environment, we provide one smooth (walking, slow motions) and one dynamic (running, aggressive motions) sequence. Ground truth positions were recorded from a Leica MS60 station with approximately 3 cm accuracy.

C. ENWIDE Results

While our approach showed improved accuracy in Table I, the main motivation behind this work is to leverage intensity to improve robustness of LIO in challenging scenarios. We therefore evaluate on the challenging ENWIDE Dataset, presented in Table II. It is plausible, that KISS-ICP fails in all sequences as it only operates on the (degenerate) point cloud geometry. However, we observe that MD-SLAM and Du, which also leverage the intensity channel, diverge in all sequences too. Both do not use the IMU, unlike LIO approaches, which impacts their ability to handle segments of geometric degeneracy or fast rotations. Additionally, Du only uses the images for geometric feature selection, and cannot benefit from additional texture information in the optimization. Due to noise in the IMU measurements and drifting biases, LIO approaches can still fail in longer segments of geometric degeneracy. This is evident in LIO-SAM, which uses curvature-based point cloud features [9]. FAST-LIO2, which operates on points directly, avoids divergence where the present vegetation still offers some weak information (*Intersection*, *Field*, *Katzensee*) but exhibits large drift. However,

TABLE II
ENWIDE DATASET - ABSOLUTE TRAJECTORY ERROR (RMSE) (*m*) / RELATIVE ERROR (%)

Method Length *(m)*	TunnelS 251.58	TunnelD 179.71	IntersectionS 279.28	IntersectionD 388.47	RunwayS 333.57	RunwayD 357.14	FieldS 232.70	FieldD 287.91	KatzenseeS 242.88	KatzenseeD 177.20
KISS-ICP [10]	× / 144.41	× / 68.11	× / 65.69	× / 64.84	× / 113.45	× / 124.64	× / 54.84	× / 70.70	× / 66.23	× / 76.80
MD-SLAM [6]	× / 88.16	× / 80.76	× / 90.87	× / 87.89	× / 97.73	× / 91.13	× / 96.03	× / 84.86	× / 93.92	× / 91.29
Du and Beltrame [8]	× / 58.084	× / 56.086	× / 60.812	× / 57.449	× / 67.906	× / 63.978	× / 72.548	× / 69.480	× / 93.92	× / 74.207
LIO-SAM [12]	× / 2565.621	× / 2662.983	× / 2022.878	× / 2362.314	× / 2334.174	× / 3984.588	× / 2196.344	× / 1999.968	5.588 / 2.673	× / 1485.377
FAST-LIO2 [1]	× / 316.12	× / 81.31	12.473 / 29.28	23.800 / 28.11	× / 53.64	× / 59.84	**0.163 / 0.57**	9.209 / 16.08	1.122 / 4.31	1.02 / 2.38
RI-LIO [7]	× / 70.32	× / 63.02	× / 49.94	× / 188.83	× / 52.18	× / 79.16	1.721 / 2.44	24.851 / 25.89	× / 49.34	× / 154.19
Ours	**0.743 / 1.60**	**0.487 / 1.59**	**0.466 / 1.25**	**1.912 / 1.69**	**1.033 / 1.89**	**2.437 / 2.98**	0.232 / 0.85	**0.581 / 1.83**	**0.412 / 0.99**	**0.592 / 1.61**

we observe divergence in environments where the geometry is effectively perfectly degenerate (*Tunnel*, *Runway*). Despite using reflectivity, RI-LIO also diverges in most sequences. By leveraging the complementary information provided by the photometric error minimization, our approach achieves robust performance in all tested sequences, While our main improvement is the increased robustness in difficult scenarios where other approaches fail, we also note higher accuracy than FAST-LIO2 on most successful sequences.

D. Ablation study

We show the effects of our image processing and feature selection in Table III. We compare the proposed *Filtered* image with *Intensity* and *Reflectivity* images. We also evaluate different feature selection policies by comparing *Random* (similar to RI-LIO [7]) as well as *Strongest* image gradient selection with the proposed geometrically *Complementary* selection. We note lower error from (*Intensity, Strongest*) than (*Reflectivity, Strongest*). This seems surprising at first, as the reflectivity value compensates the range dependency of the signal. However, we observed that the reflectivity image contains stronger noise and artefacts and has less consistent brightness across the image. Our proposed image processing (*Filtered, Strongest*) improves performance in low-textured environments (*TunnelD, KatzenseeD*), where the line artefacts are more dominant than the actual features from the environment. Additionally, the brightness decreases drastically with increasing range. In contrast, the brightness compensation and line removal of our filtered image allow us to use more fine-grained details, e.g. from vegetation or gravel. We note that (*Intensity, Strongest*) marginally outperforms (*Filtered, Strongest*) on *IntersectionS*. In this scene, strong image features from road cracks are consistently found at short range. We believe that the slightly lower ATE results from the filtered image having lower contrast than the intensity image at short range in this scene, which results in weaker gradients. Selecting features based on strong image gradients (*Filtered, Strong*) results in better performance compared to random patches (*Filtered, Random*), as they provide richer information. Our proposed feature selection scheme (*Filtered, Complementary*) achieves the best performance, as it reduces redundant information along uninformative geometric directions and specifically selects informative image patches. The impact is strongest in *TunnelD*, where most gradients are in the geometrically degenerate direction along the tunnel (see Figure 5), while they are more randomly oriented in the other scenes. Overall, the ablation experiments confirm that COIN-LIO is able to effectively leverage the additional information provided by the multi-modality of the approach.

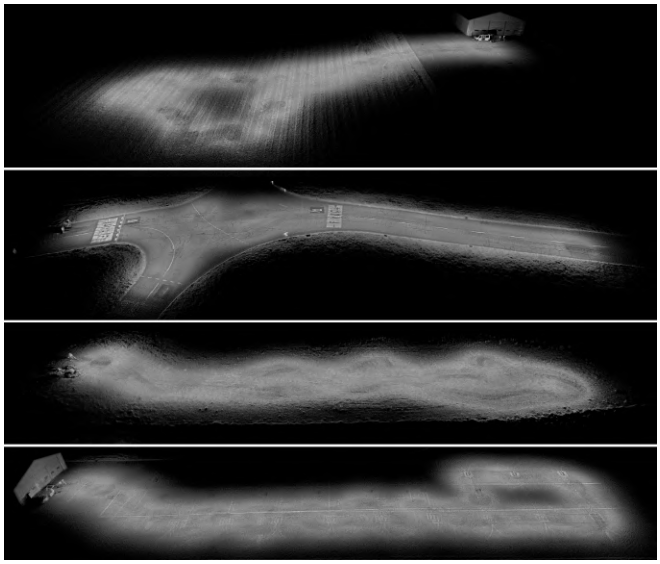

Fig. 6. Resulting maps from COIN-LIO on the ENWIDE dataset. Top to bottom: FieldS, IntersectionS, KatzenseeS, RunwayS. Despite long degenerate sections, COIN-LIO produces consistent, sharp maps.

TABLE III
ABLATION STUDY - ABSOLUTE TRAJECTORY ERROR (RMSE) (*m*)

Image	Features	TunnelD	IntersectionS	KatzenseeD
Intensity	Strongest	13.928	0.472	0.948
Reflectivity	Strongest	×	0.699	0.874
Filtered	Strongest	0.814	0.489	0.701
Filtered	Random	1.913	0.580	1.073
Filtered	Complementary	**0.487**	**0.466**	**0.592**

V. CONCLUSION

We proposed COIN-LIO, a LiDAR-inertial odometry framework that fuses photometric error minimization on LiDAR intensity images with geometric registration to improve robustness in geometrically degenerate environments. We presented a filtering pipeline to produce brightness-compensated intensity images that provide more details and consistent illumination across different scenes. Our novel feature selection scheme effectively leverages the multi-modality by providing additional instead of redundant information. While COIN-LIO requires high-resolution LiDARs for dense intensity images, it slightly outperforms baseline approaches on the geometry-rich Newer College Dataset and shows drastically increased robustness in our new, geometrically degenerate ENWIDE dataset, which enables benchmarking in previously underrepresented scenarios.

We believe that this dataset as well as our work serve as a motivation for a new line of research that shifts from chasing even higher accuracy in geometrically simple cases to improving robustness in challenging environments. We also hope it motivates the industry to further improve the imaging capabilities of LiDAR.

REFERENCES

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

[2] J. Zhang and S. Singh, "Laser–visual–inertial odometry and mapping with high robustness and low drift," *Journal of field robotics*, vol. 35, no. 8, pp. 1242–1264, 2018.

[3] X. Zuo *et al.*, "Lic-fusion 2.0: Lidar-inertial-camera odometry with sliding-window plane-feature tracking," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 5112–5119.

[4] D. Wisth, M. Camurri, and M. Fallon, "Vilens: Visual, inertial, lidar, and leg odometry for all-terrain legged robots," *IEEE Transactions on Robotics*, 2022.

[5] 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.

[6] L. Di Giammarino, L. Brizi, T. Guadagnino, C. Stachniss, and G. Grisetti, "Md-slam: Multi-cue direct slam," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 11 047–11 054.

[7] Y. Zhang *et al.*, "Ri-lio: Reflectivity image assisted tightly-coupled lidar-inertial odometry," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1802–1809, 2023.

[8] W. Du and G. Beltrame, "Real-time simultaneous localization and mapping with lidar intensity," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 4164–4170.

[9] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time.," in *Robotics: Science and systems*, Berkeley, CA, vol. 2, 2014, pp. 1–9.

[10] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss, "Kiss-icp: In defense of point-to-point icp – simple, accurate, and robust registration if done the right way," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 1029–1036, 2023.

[11] T. Tuna, J. Nubert, Y. Nava, S. Khattak, and M. Hutter, *X-icp: Localizability-aware lidar registration for robust localization in extreme environments*, 2023. arXiv: 2211.16335 [cs.RO].

[12] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5135–5142.

[13] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "Isam2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 3281–3288.

[14] F. Dellaert and M. Kaess, "Factor graphs for robot perception," *Found. Trends Robotics*, vol. 6, pp. 1–139, 2017.

[15] W. Xu and F. Zhang, "Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3317–3324, 2021.

[16] Y. S. Park, H. Jang, and A. Kim, "I-loam: Intensity enhanced lidar odometry and mapping," in *2020 17th International Conference on Ubiquitous Robots (UR)*, 2020, pp. 455–458.

[17] H. Wang, C. Wang, and L. Xie, "Intensity-slam: Intensity assisted localization and mapping for large scale environment," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1715–1721, 2021.

[18] H. Li, B. Tian, H. Shen, and J. Lu, "An intensity-augmented lidar-inertial slam for solid-state lidars in degenerated environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[19] S. Li, B. Tian, X. Zhu, J. Gui, W. Yao, and G. Li, "Inten-loam: Intensity and temporal enhanced lidar odometry and mapping," *Remote Sensing*, vol. 15, no. 1, 2023, ISSN: 2072-4292.

[20] Y. Pan, P. Xiao, Y. He, Z. Shao, and Z. Li, "Mulls: Versatile lidar slam via multi-metric linear least square," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 11 633–11 640.

[21] C. McManus, P. Furgale, and T. D. Barfoot, "Towards lighting-invariant visual navigation: An appearance-based approach using scanning laser-rangefinders," *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 836–852, 2013.

[22] T. D. Barfoot *et al.*, "Into darkness: Visual navigation based on a lidar-intensity-image pipeline," in *Robotics Research: The 16th International Symposium ISRR*, Springer, 2016, pp. 487–504.

[23] C. McManus, P. Furgale, and T. D. Barfoot, "Towards appearance-based methods for lidar sensors," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 1930–1935. DOI: 10.1109/ICRA.2011.5980098.

[24] H. Dong and T. D. Barfoot, "Lighting-invariant visual odometry using lidar intensity imagery and pose interpolation," in *Field and Service Robotics: Results of the 8th International Conference*, Springer, 2013, pp. 327–342.

[25] T. Guadagnino, X. Chen, M. Sodano, J. Behley, G. Grisetti, and C. Stachniss, "Fast sparse lidar odometry using self-supervised feature selection on intensity images," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7597–7604, 2022.

[26] D. He, W. Xu, and F. Zhang, "Kalman filters on differentiable manifolds," *arXiv preprint arXiv:2102.03804*, 2021.

[27] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "Svo: Semidirect visual odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2017.

[28] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[29] C. Harris, M. Stephens, *et al.*, "A combined corner and edge detector," in *Alvey vision conference*, Citeseer, vol. 15, 1988, pp. 10–5244.

[30] L. Zhang, M. Camurri, D. Wisth, and M. Fallon, "Multi-camera lidar inertial extension to the newer college dataset," *arXiv preprint arXiv:2112.08854*, 2021.