PI-EAT-SLAM: Physics-Informed Edge-Aware Thermal Gaussian Splatting SLAM

Junpyo Kim[1], Seonwook Yeom[1], Seonmo Yang[1], Jungil Ham[1]
Hyeonji Shin[1], Changhyeon Kim[2], Hyeonbeom Lee[3], and Pyojin Kim[1]

***Abstract*— Conventional SLAM methods relying on photometric alignment or point features often fail in texture-sparse, low-contrast thermal imagery. To address these issues, we present Physics-Informed Edge-Aware Thermal Gaussian Splatting SLAM (PI-EAT-SLAM), the first thermal-depth GS-SLAM framework. PI-EAT-SLAM achieves reliable tracking and mapping in visually degraded environments by exploiting robust geometric edge features rather than photometric intensity. Our approach leverages edge-based optimization by introducing a Stefan-Boltzmann law-based thermal rescaling (S-B rescaling) module to physically enhance contrast. Building on this, we employ a multi-stage Gradient-aware Edge-KLT (GE-KLT) tracking method with rigorous dual-constraint (spatial and gradient-directional) outlier rejection for highly reliable correspondences. Finally, the mapping process integrates an edge-aware smoothness loss that preserves sharp thermal boundaries while enforcing thermal equilibrium. Extensive experiments demonstrate that PI-EAT-SLAM achieves superior tracking accuracy and highly competitive novel view synthesis performance compared to state-of-the-art baselines. The code is available on the project page: `https://pi-eat-slam.github.io`**

Fig. 1. Overview of PI-EAT-SLAM on the custom dataset, demonstrating robust tracking and mapping in low-contrast, texture-sparse thermal imagery. (Background) Overall visualization of the 3DGS mapping and camera trajectory. (Red box) Robust edge point matching for tracking: red and blue dots denote edge points extracted from the rendered and ground-truth images, respectively, with cyan lines indicating the residual vectors.

I. Introduction

Robust scene understanding in visually degraded conditions is critical for autonomous systems. Conventional visual SLAM and recent 3D Gaussian Splatting (3DGS)-based RGB/RGB-D SLAM [1], [2] methods perform poorly in these scenarios, as they rely heavily on dense visual textures and photometric consistency. Thermal imaging offers an illumination-independent alternative, but inherently suffers from low contrast, sparse textures, and high sensor noise, complicating standard photometric alignment. Existing solutions often employ multi-sensor fusion [3], [4], [5], which significantly increases calibration complexity or hardware payload. Meanwhile, recent implicit thermal reconstruction methods [6] struggle to scale beyond object-level mapping, leaving thermal 3DGS SLAM largely unexplored.

To address these limitations, we propose PI-EAT-SLAM (Physics-Informed Edge-Aware Thermal Gaussian Splatting SLAM) for lightweight thermal-depth inputs. We introduce a Stefan-Boltzmann (S-B) rescaling module to physically enhance thermal contrast without amplifying noise. This critical preprocessing empowers us to replace failure-prone photometric alignment with robust edge-based tracking and mapping. Evaluations on public [7] and custom datasets demonstrate overwhelming tracking superiority and competitive novel view synthesis. By operating efficiently where conventional methods fail, our approach opens a new horizon for 3D GS-based SLAM in the thermal domain.

Our main contributions are summarized as follows:

- A physics-informed thermal preprocessing method with the Stefan-Boltzmann law, which optimally enhances contrast while preserving radiometric consistency.
- GE-KLT tracking, which robustly extracts edge correspondences under severe thermal noise via a multi-stage pipeline and rigorous dual-constraint outlier rejection.
- An edge-aware smoothness loss for 3DGS mapping that enforces natural thermal equilibrium while adaptively preserving sharp structural discontinuities.
- Extensive experimental validation on public and custom datasets, demonstrating significant performance improvements over state-of-the-art baselines.

II. Related Work

Thermal Image Rescaling. Converting 14/16-bit thermal data to an 8-bit format is a critical preprocessing step for thermal SLAM, as improper scaling inherently causes lower contrast, amplified noise, and photometric inconsistency.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2024-00358374) and Global University Project grant funded by the GIST in 2025. *(Corresponding author: Pyojin Kim)*

[1]Department of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea. `junpyokim@gm.gist.ac.kr`, `pjinkim@gist.ac.kr`

[2]Holiday Robotics, Seoul 06234, South Korea. `changhyeon.kim@holiday-robotics.com`

[3]Department of Mechanical Engineering, Ajou University, Suwon 16499, South Korea. `hbeomlee@ajou.ac.kr`

Min-max scaling utilized for raw thermal images [8] fails to adequately enhance low contrast, while adaptive image enhancement or online photometric calibration methods (e.g., [9], [10]) often amplify noise magnitudes or fail to fully restore true contrast. This severely degrades Gaussian Splatting SLAM (GS-SLAM) tracking and rendering.

Thermal Odometry and SLAM. Existing thermal SLAM often relies on multi-sensor fusion to address low-texture and high-noise characteristics. However, incorporating an IMU [3], [4] requires strict time synchronization and complex calibration, whereas LiDAR fusion [5] additionally imposes a severe hardware payload burden due to the sensor's heavy weight. While a recent thermal-depth approach [11] attempts to mitigate these issues by utilizing raw thermographic data and a semi-direct framework to handle camera Non-Uniformity Correction (NUC) interruptions, it still relies fundamentally on direct image alignment. Consequently, it lacks explicit geometric constraints and struggles heavily in severely low-texture or thermally-flat regions. Furthermore, implicit neural representations [6] struggle to scale beyond object-level mapping.

Gaussian Splatting-Based SLAM. While 3D GS enables high-fidelity SLAM [1], [2], its extension to the thermal domain remains largely unexplored. Even a state-of-the-art framework that incorporates visual geometric priors [12] fundamentally relies on photometric losses, which tend to degrade under the high-noise, low-contrast thermal imagery.

Edge-Based Visual Odometry/SLAM. Edge features provide robustness to illumination changes [13], [14], [15]. Nevertheless, existing methods struggle in feature-sparse thermal images because they establish correspondences based solely on spatial proximity and gradient consistency while completely ignoring local patch appearances [13], [14], or because they fail to verify tracked points against high-confidence structural edges [15]. In contrast, we explicitly exploit local patch intensities via KLT tracking [16] and rigorously filter reliable matches from Canny edges [17] using a dual-constraint (spatial and gradient) outlier rejection, significantly improving tracking robustness in visually degraded scenes.

III. Preliminary

A. Notations and Camera Model

Let $\{W\}$ and $\{C\}$ denote the world and camera frames. A 3D point $\mathbf{P}_W \in \mathbb{R}^3$ is transformed to $\{C\}$ via $\mathcal{T}_{CW} = [\mathbf{R}_{CW}|\mathbf{t}_{CW}] \in \mathrm{SE}(3)$. Under the pinhole model, $\mathbf{P}_W$ projects to pixel $\mathbf{p} = (u, v)^\top$ using intrinsics $\mathbf{K} \in \mathbb{R}^{3\times 3}$:

$$Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}\left(\mathbf{R}_{CW}\mathbf{P}_W + \mathbf{t}_{CW}\right), \tag{1}$$

where Z is the point depth. Let $I(\mathbf{p})$ and $\tilde{D}(\mathbf{p})$ denote the thermal intensity and observed depth at $\mathbf{p}$, respectively.

B. Principles of Thermal Imagery

Thermal infrared cameras measure emitted radiation. By the Stefan-Boltzmann law [18], surface radiance L is proportional to the fourth power of its absolute temperature T:

$$L = \epsilon\sigma T^4, \tag{2}$$

where σ is the Stefan-Boltzmann constant and ϵ is emissivity. Raw thermal images typically suffer from low contrast and sparse textures due to marginal temperature variances. To mitigate this, we exploit the non-linear T^4 relationship to physically amplify subtle thermal differences. This physics-informed perspective provides a robust, high-contrast foundation for our edge-aware tracking framework, which we detail in the subsequent methodology.

C. 3D Gaussian Splatting Formulation

Our mapping and rendering modules build upon 3D Gaussian Splatting (3DGS). A scene is represented by anisotropic 3D Gaussians, where each Gaussian $\mathcal{G}^i$ is parameterized by its world-frame mean $\mu_W^i \in \mathbb{R}^3$, covariance Σ_W^i, opacity $\alpha^i \in [0, 1]$, and thermal radiance feature r^i. To render an image from pose $\mathcal{T}_{CW}$, a pixel intensity $I(\mathbf{p})$ is computed via front-to-back α-blending of $\mathcal{N}$ ordered Gaussians:

$$I(\mathbf{p}) = \sum_{i\in\mathcal{N}} r^i \alpha^i \prod_{j=1}^{i-1}(1 - \alpha^j), \tag{3}$$

where α^i is the 2D projected opacity. The depth map $D(\mathbf{p})$ is rasterized similarly by substituting r^i with z_i, the distance to μ_W^i along the camera ray. This differentiable rendering process drives our scene geometry optimization.

IV. Proposed Method

A. Physics-Informed Thermal Image Preprocessing

Thermal sensors typically capture data in 14-bit or 16-bit formats, which must be converted to 8-bit for standard tracking pipelines. While conventional linear min-max normalization on raw intensities [8] often loses critical thermal details, we propose a physics-informed Stefan-Boltzmann (S-B) rescaling method to robustly preserve them. First, using the camera manufacturer's formula, we convert the intensity at pixel (x, y) in the raw thermal image into an absolute temperature $T(x, y)$ (in Kelvin). By following the Stefan-Boltzmann law, we compute the thermal radiance $L(x, y)$, which is directly proportional to the fourth power of T:

$$L(x, y) = \epsilon\sigma T(x, y)^4, \tag{4}$$

To utilize this radiance map for feature tracking, we convert it into a standard 8-bit image format using a radiance-based min-max normalization, yielding the scaled intensity $I_{\text{scaled}}(x, y)$:

$$I_{\text{scaled}}(x, y) = \frac{L(x, y) - L_{\min}}{L_{\max} - L_{\min}} \times 255. \tag{5}$$

Constant scaling factors, such as the Stefan-Boltzmann constant σ and the assumed uniform scene emissivity ϵ, naturally cancel out during this normalization process, meaning exact emissivity values are not required. Compared to the conventional approach of applying linear min-max scaling directly to the raw thermal images [8], our proposed

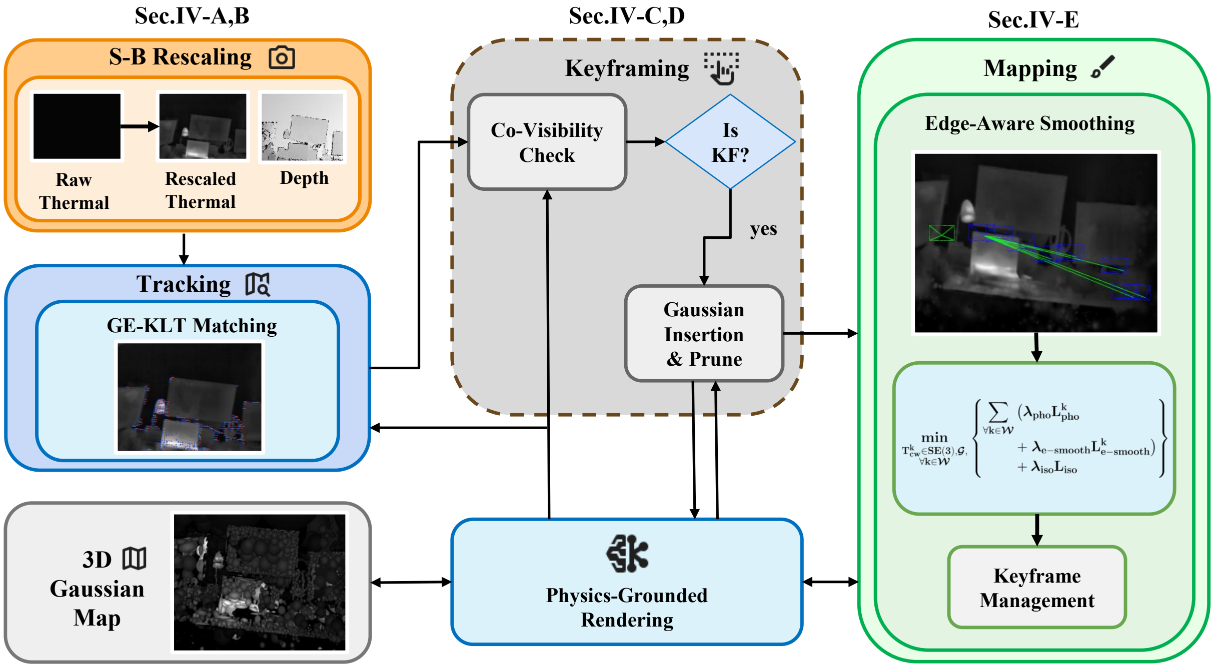


Fig. 2. **Overview of the PI-EAT-SLAM Pipeline.** (Left) Stefan–Boltzmann (S-B) rescaling enhances thermal contrast to enable reliable edge extraction. This drives our Gradient-aware Edge-KLT (GE-KLT) tracking, achieving robust pose estimation via a multi-stage matching pipeline (KLT → nearest-neighbor) with rigorous dual-constraint (spatial and gradient) outlier rejection. (Center) It manages keyframes and Gaussian updates, utilizing physics-grounded rendering [19] with an atmospheric transmission field (ATF) to suppress ghosting artifacts and a thermal conduction module (TCM) for edge-enhanced rendering. (Right) An edge-aware smoothness loss is applied during 3DGS mapping to ensure high-fidelity dense reconstruction.

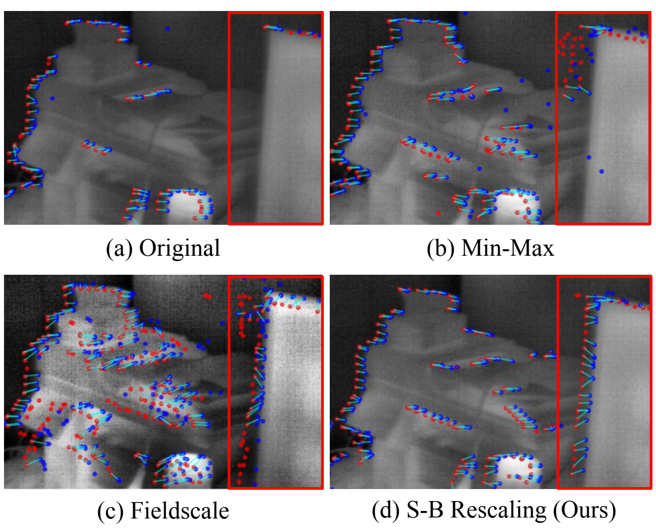


Fig. 3. Comparison of edge point matching results across different rescaling methods: (a) original thermal image from the Multi-spectral dataset [7], (b) min-max scaling [8], (c) Fieldscale [9], and (d) our Stefan-Boltzmann rescaling. Red and blue dots represent detected edge points on the rendered and GT images, respectively, and cyan lines indicate the final established correspondences. Our method achieves the most reliable and consistent matches by effectively enhancing contrast while suppressing noise.

method yields an image that much more effectively reflects true physical thermal radiation. This non-linear mapping (T^4) inherently amplifies subtle but physically meaningful temperature variations, creating a sharp 8-bit image.

As visualized in Fig. 3, and as seen in the red-boxed region, our physics-informed preprocessing significantly improves edge matching. The original thermal image yields very few matches due to its inherent low contrast. Conventional min-max scaling [8] fails to accurately reflect thermal characteristics and maximize contrast due to its linear rescaling process, leading to numerous incorrect correspondences. While FieldScale [9] significantly increases contrast, it severely amplifies sensor noise, generating widespread false matches. Conversely, our Stefan-Boltzmann rescaling optimally enhances contrast without amplifying noise, extracting highly accurate matches along physical boundaries. Finally, following [20], we apply CLAHE [21] and a bilateral filter [22] to further enhance local contrast and attenuate noise. As shown in Fig. 4, this complete pipeline preserves sharp boundaries while smoothing homogeneous regions, establishing an optimal foundation for edge-based tracking.

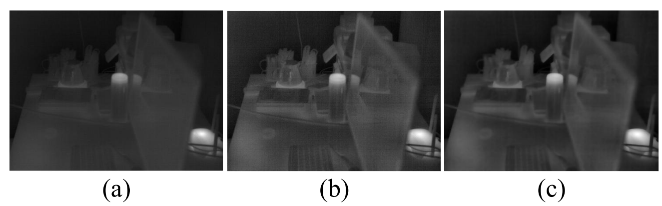


Fig. 4. Step-by-step thermal image preprocessing results: (a) our S-B rescaling for global contrast enhancement, (b) S-B rescaling + CLAHE for local detail visibility, and (c) S-B rescaling + CLAHE + Bilateral filtering (final output) to preserve edge sharpness while reducing noise.

B. GE-KLT Tracking

As illustrated in Fig. 5, our multi-stage matching pipeline combines KLT tracking [16] (leveraging local patch appearance) with high-confidence Canny edge points [17]. Specifically, we follow three sequential outlier rejection stages:

(1) distance-based filtering, (2) dual-constraint validation, and (3) residual-based rejection. Through these stages, we ensure that only reliable correspondences contribute to pose optimization, significantly improving tracking robustness in contrast-limited thermal images.

1) KLT Tracking with Distance-Based Outlier Rejection: We extract Canny edge points from rendered and ground-truth (GT) frames. Each rendered edge point p_i^{rend} is tracked to a point p_i^{track} in the GT frame via KLT. The track is valid only if the Euclidean distance is below a threshold τ_{dist}:

$$\left\|p_i^{\text{track}} - p_i^{\text{rend}}\right\| < \tau_{\text{dist}}. \tag{6}$$

2) Dual-Constraint Edge Matching: Each valid point p_i^{track} is matched to its nearest GT Canny edge point p_i^{gt}. We compute the normalized image gradient $\mathbf{n}(p)$ via Sobel filters:

$$\mathbf{n}(p) = \frac{\nabla I(p)}{|\nabla I(p)|}. \tag{7}$$

A match is accepted only if it satisfies both spatial and gradient-directional constraints:

$$\|p_i^{\text{gt}} - p_i^{\text{track}}\| < \tau_{\text{dist}}, \tag{8}$$

$$\theta_i = \arccos(\mathbf{n}_i^{\text{track}} \cdot \mathbf{n}_i^{\text{gt}}) < \tau_{\text{angle}}, \tag{9}$$

where θ_i is the angle between the gradient vectors. Setting $\tau_{\text{dist}} = 25$ pixels and $\tau_{\text{angle}} = 45°$ empirically ensures robust rejection of implausible matches.

3) Residual and Loss Formulation: For each matched pair $(p_i^{\text{rend}}, p_i^{\text{gt}})$, the residual r_i couples spatial distance with gradient angular difference:

$$r_i = \left\|p_i^{\text{gt}} - p_i^{\text{rend}}\right\| \cdot \arccos\left(\mathbf{n}(p_i^{\text{rend}}) \cdot \mathbf{n}(p_i^{\text{gt}})\right). \tag{10}$$

This formulation, adopted from [15] within our novel multi-stage pipeline, simultaneously minimizes spatial distances and enforces rotational alignment of the corresponding edgelets. After discarding matches exceeding τ_{dist}, the Gradient-aware Edge-KLT (GE-KLT) loss is defined as:

$$L_{\text{GE-KLT}} = \frac{1}{N_{\text{inlier}}} \sum_{i \in \mathcal{I}} r_i^2, \tag{11}$$

where $\mathcal{I}$ is the set of inliers and $N_{\text{inlier}} = |\mathcal{I}|$.

Complementary to GE-KLT, we apply an L1 depth loss against the observed GT depth $\tilde{D}$:

$$L_{\text{geo}} = \left\|D(\mathcal{G}, \mathcal{T}_{CW}) - \tilde{D}\right\|_1, \tag{12}$$

where $D(\mathcal{G}, \mathcal{T}_{CW})$ is the depth rendered from 3D Gaussians $\mathcal{G}$ at pose $\mathcal{T}_{CW}$. We only consider valid depth pixels (depth > 0.01 and opacity > 0.95). The overall tracking loss is:

$$L_{\text{tracking}} = \lambda_{\text{GE-KLT}} L_{\text{GE-KLT}} + \lambda_{\text{geo}} L_{\text{geo}}, \tag{13}$$

where $\lambda_{\text{GE-KLT}}$ and λ_{geo} are empirical weighting coefficients that balance the edge and geometric constraints.

C. Keyframing

We follow the keyframing strategy from [1], which sequentially manages co-visibility assessment, keyframe selection, and Gaussian refinement.

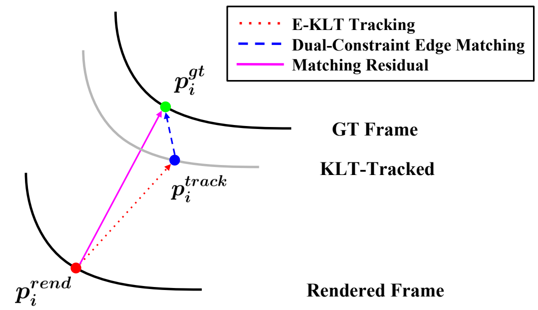


Fig. 5. Illustration of the GE-KLT residual. The Rendered Frame denotes the region where the edge points (p_i^{rend}) detected in the rendered image are located. 'KLT-Tracked' represents the region where the tracked points (p_i^{track}) are located, which are obtained by tracking p_i^{rend} toward the ground-truth (GT) image using E-KLT tracking (red dotted line). The GT Frame indicates the region where the final edge points (p_i^{gt}) are located, which are matched with p_i^{track} among the detected GT edges via our dual-constraint edge matching (blue dashed line). The solid magenta line represents the resulting matching residual.

D. Rendering

To enhance rendering fidelity without additional optimization overhead, we integrate the pre-trained Atmospheric Transmission Field (ATF) and Thermal Conduction Module (TCM) from [19], [23]. The ATF mitigates floating artifacts by adjusting the original spherical harmonics SH_0 with learned absorption (μ_{abs}) and scattering (μ_{sca}) coefficients:

$$SH = SH_0 e^{(\mu_{abs} + \mu_{sca})d}, \tag{14}$$

where d is the propagation distance. Concurrently, the TCM sharpens blurred boundaries using 2D thermal conduction:

$$\frac{\partial u}{\partial t} = \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \tag{15}$$

where u and α represent the temperature field and thermal conduction rate, respectively.

E. Mapping

Conventional smoothness losses [24] tend to oversmooth important geometric boundaries. To suppress spurious radiance oscillations while explicitly preserving edge structures, we propose an edge-aware smoothness loss:

$$L_{\text{e-smooth}} = \frac{1}{4M} \sum_{i,j} \Big(|R_{i\pm1,j} - R_{i,j}| + |R_{i,j\pm1} - R_{i,j}|\Big) \cdot (1 - w_{\text{edge}}(i,j)), \tag{16}$$

where $R_{i,j}$ is the rendered thermal radiance at pixel (i,j) after rescaling, and M is the total pixel count. The edge weight $w_{\text{edge}}(i,j)$ is defined as:

$$w_{\text{edge}}(i,j) = \sigma\big(\alpha(|\nabla R_{i,j}| - \beta)\big), \tag{17}$$

where $|\nabla R_{i,j}|$ is the Sobel gradient magnitude and $\sigma(x) = 1/(1+e^{-x})$ is the sigmoid function. With empirical parameters $\alpha = 8$ (sensitivity) and $\beta = 0.1$ (threshold), this formulation enforces strong smoothing in flat regions ($w_{\text{edge}} \approx 0$) while strictly preserving boundaries near strong edges ($w_{\text{edge}} \approx 1$).

TABLE I
CAMERA TRACKING ACCURACY (ATE RMSE IN METERS) ON MULTI-SPECTRAL DATASET. BEST: **BOLD**, SECOND BEST: UNDERLINED.

Sequence		ORB2 (RGB-D)	ORB2 (Thermal-D)	MonoGS (Thermal-D)	Thermal-D Odom	Ours
Bright	ds1-xyz	**0.012**	0.083	0.064	0.072	0.054
	ds1-xyz-ps	0.040	0.141	0.042	**0.015**	0.031
	ds1-hfsp	0.106	0.442	0.096	0.058	**0.049**
	ds2-xyz	0.062	Fail	Fail	**0.038**	0.087
	ds2-hfsp	0.076	0.153	0.154	**0.029**	0.144
Varying	ds1-xyz-ic	Fail	0.250	0.082	0.115	**0.055**
	ds1-xyz-ic-ps	Fail	0.349	0.061	0.078	**0.052**
	ds1-hfsp-ic	Fail	0.467	0.080	0.051	**0.045**
	ds2-xyz-ic	Fail	0.203	0.171	0.114	**0.084**
	ds2-hfsp-ic	Fail	0.189	0.099	0.110	**0.063**

The final mapping loss aggregates the photometric loss L^k_{pho}, our edge-aware smoothness loss $L^k_{\text{e-smooth}}$ over keyframes k in window W, and isotropic regularization L_{iso}:

$$L_{\text{mapping}} = \textstyle\sum_{k \in W} \left(\lambda_{\text{pho}} L^k_{\text{pho}} + \lambda_{\text{e-smooth}} L^k_{\text{e-smooth}} \right) + \lambda_{\text{iso}} L_{\text{iso}}, \quad (18)$$

where λ_{pho}, $\lambda_{\text{e-smooth}}$, and λ_{iso} are empirical weighting coefficients that balance rendering fidelity, edge preservation, and geometric regularization.

V. EVALUATION

We evaluate our method on two distinct datasets to demonstrate its robustness across different sensor configurations.

Multi-spectral Dataset. We utilize the public Multi-spectral dataset [7] featuring hardware-synchronized thermal and RGB-D sequences with 120 Hz OptiTrack ground truth. We evaluate tracking, novel view synthesis (NVS), and trajectory consistency. Sequences are categorized into *Bright* (stable lighting) and *Varying* (random light toggling).

Custom Monocular Thermal Dataset. We collect a custom thermal dataset using the thermal camera of a DJI Mavic 3T drone, with ground-truth poses captured by an OptiTrack motion capture system. The sequences feature *Easy* (translation-dominant, feature-rich) and *Difficult* (rotation-dominant, feature-sparse) trajectories. Although PI-EAT-SLAM targets Thermal-Depth setups, where missing depth inherently degrades monocular rendering and edge matching, we evaluate it in a purely monocular setting to validate our tracking pipeline due to the absence of open-source Thermal-Depth baselines. For fair comparison, we benchmark against TI-SLAM [4] (adapted without IMU), MonoGS [1], HI-SLAM2 [12], and ORB-SLAM3 [25].

A. Multi-spectral Dataset

For brevity in the evaluation tables, sequence names are abbreviated (e.g., *desk* to *ds*, *person* to *ps*, *halfsphere* to *hfsp*, and *illumination changes* to *ic*).

1) Tracking Performance: Table I compares our tracking performance against ORB-SLAM2 [26], MonoGS [1], and a state-of-the-art Thermal-Depth Odometry [11].

Thermal-Depth Odometry relies on semi-direct photometric patch alignment, which degrades significantly when patch-based tracking fails in low-contrast thermal regions. Similarly, ORB-SLAM2 (Thermal-D) yields large errors, confirming that discrete point matching is inadequate for thermal domains. ORB-SLAM2 (RGB-D) fails completely in all *Varying* sequences due to aggressive illumination changes. MonoGS demonstrates competitive performance in some scenes but relies purely on photometric loss, causing it to completely fail tracking and initialization in the dark *ds2-xyz* sequence. Overall, PI-EAT-SLAM achieves superior accuracy across most sequences, particularly in challenging *Varying* conditions. These results validate that our explicit geometric edge-matching strategy provides significantly more robust localization than methods relying on discrete patches or photometric optimization.

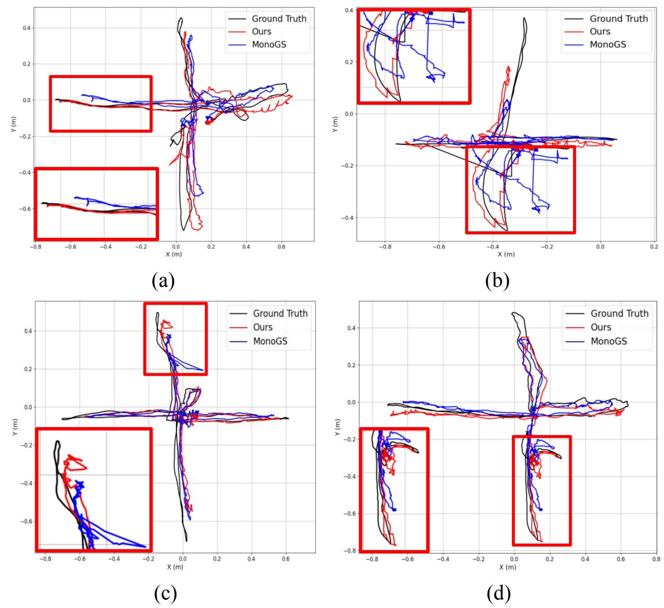


Fig. 6. Trajectory comparison on the multi-spectral dataset. PI-EAT-SLAM (Red) and MonoGS [1] (Blue) are compared against Ground Truth (Black). The evaluated sequences are (a) desk1-halfsphere, (b) desk2-xyz-ic, (c) desk1-xyz-ic-person, and (d) desk1-halfsphere-ic. Red boxes highlight segments where MonoGS significantly deviates, whereas our method maintains accurate alignment, demonstrating superior tracking robustness.

2) Rendering Quality: Table II presents quantitative NVS metrics. MonoGS naturally scores higher by directly minimizing photometric loss. Conversely, our preprocessing intentionally alters intensity distributions to prioritize tracking in low-texture environments, resulting in a numerical NVS trade-off. However, this design is critical: it enables stable tracking in dark sequences (e.g., *ds2-xyz*) where MonoGS completely fails to initialize Gaussians. As Thermal SLAM-NeRF [6] is not open-source, we report the available PSNR values directly from their paper and denote the unreported SSIM and LPIPS metrics as '-'.

Furthermore, standard metrics fail to adequately capture structural fidelity. As shown in Fig. 7, MonoGS blurs edges and loses details like laptop keys. In contrast, our edge-aware approach preserves sharp boundaries and fine geometric structures, confirming its superiority in both tracking robustness and 3D structural reconstruction.

B. Custom Monocular Thermal Dataset

1) Tracking Performance: Table III reports the tracking evaluation. Traditional feature-based methods like ORB-

TABLE II
NOVEL VIEW SYNTHESIS PERFORMANCE ON MULTI-SPECTRAL DATASET. BEST RESULTS ARE IN **BOLD**.

		Bright					Varying				
Methods	**Metrics**	**ds1-xyz**	**ds1-xyz-ps**	**ds1-hfsp**	**ds2-xyz**	**ds2-hfsp**	**ds1-xyz-ic**	**ds1-xyz-ic-ps**	**ds1-hfsp-ic**	**ds2-xyz-ic**	**ds2-hfsp-ic**
MonoGS (Thermal-D)	PSNR↑	**36.11**	**35.69**	**36.74**	Fail	22.12	**29.83**	**33.37**	**36.17**	26.02	**31.46**
	SSIM↑	**0.922**	**0.927**	**0.946**	Fail	0.808	**0.813**	**0.863**	**0.942**	**0.832**	**0.878**
	LPIPS↓	**0.180**	**0.159**	**0.202**	Fail	**0.518**	**0.293**	**0.280**	**0.222**	**0.501**	**0.474**
Thermal SLAM -NeRF (Mono) [6]	PSNR↑	30.23	-	-	-	**27.45**	29.53	-	29.09	**29.61**	-
	SSIM↑	-	-	-	-	-	-	-	-	-	-
	LPIPS↓	-	-	-	-	-	-	-	-	-	-
Ours	PSNR↑	29.85	28.73	31.26	**21.75**	22.06	28.18	26.12	32.84	24.66	27.77
	SSIM↑	0.881	0.889	0.918	**0.795**	**0.819**	0.783	0.828	0.920	0.800	0.847
	LPIPS↓	0.277	0.239	0.248	**0.472**	0.572	0.322	0.396	0.243	0.561	0.538

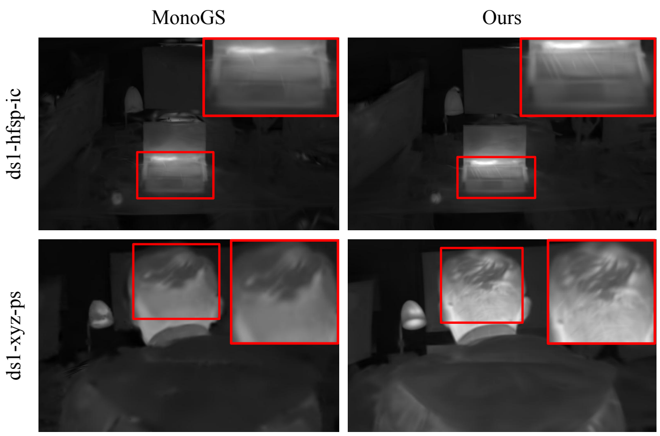


Fig. 7. Qualitative 3D map comparison on the *desk1-halfsphere-ic* and *desk1-xyz-person* sequences from the Multi-spectral dataset. MonoGS [1] blurs edges and yields inaccurate reconstruction, collapsing local details such as the laptop keypad and the fine details of the person's hair (red boxes). Ours preserves sharp structural boundaries and fine geometric details.

TABLE III
ATE RMSE (M) ON CUSTOM DATASET (MONOCULAR). BEST: **BOLD**, SECOND BEST: UNDERLINED.

Difficulty	**Sequence**	**ORB-SLAM3**	**MonoGS**	**HI-SLAM2**	**TI-SLAM (Mono)**	**Ours (Mono)**
Easy	xyz	**0.1081**	0.1339	0.1816	0.6278	0.1270
	sfm	0.1063	0.0874	0.1217	1.0422	**0.0870**
	circle	**0.1187**	0.2917	0.2117	1.0446	0.1484
Difficult	upper_turn	Fail	0.9159	0.8996	0.9212	**0.8892**
	down_turn	Fail	0.8313	0.8241	0.8530	**0.8234**
	up_down	Fail	**1.1383**	1.1874	1.2361	1.1852

SLAM3 [25] completely fail on the 'Difficult' sequences due to aggressive motions and texture-less thermal images. MonoGS demonstrates reasonable tracking but, constrained by its pure photometric loss, falls behind our edge-optimized approach. This limitation is particularly evident in continuous circular maneuvers (e.g., *circle*), where MonoGS struggles to maintain precise localization, yielding an ATE RMSE nearly twice ours. We also adapted TI-SLAM [4] as a thermal baseline by removing its IMU input and retraining it for monocular thermal inference. It exhibits large ATE RMSEs because its pure deep regression architecture, lacking explicit geometric modeling, fails to generalize to feature-sparse

TABLE IV
NOVEL VIEW SYNTHESIS ON CUSTOM DATASET (MONOCULAR). BEST: **BOLD**, SECOND BEST: UNDERLINED.

Difficulty		**Easy**			**Difficult**		
Methods	**Metrics**	**xyz**	**sfm**	**circle**	**upper turn**	**down turn**	**up down**
MonoGS	PSNR↑	**25.54**	27.22	21.81	15.31	11.36	12.66
	SSIM↑	0.941	0.943	0.876	0.528	0.268	0.284
	LPIPS↓	0.179	0.145	0.207	0.540	0.690	0.656
HI-SLAM2	PSNR↑	23.66	24.41	22.36	19.43	**18.54**	**20.10**
	SSIM↑	0.916	0.895	0.885	0.855	**0.846**	**0.865**
	LPIPS↓	0.193	0.196	0.207	0.464	**0.467**	**0.400**
Ours (Mono)	PSNR↑	25.42	**27.29**	**23.73**	**20.82**	11.85	15.26
	SSIM↑	**0.944**	**0.944**	**0.920**	**0.870**	0.326	0.436
	LPIPS↓	**0.168**	**0.141**	**0.158**	**0.378**	0.661	0.574

scenes. Furthermore, our method consistently outperforms HI-SLAM2 [12], a state-of-the-art RGB GS-SLAM framework. This validates that integrating edge-based geometric constraints is essential for overcoming the unique challenges of the thermal domain.

2) Rendering Quality: As shown in Table IV, our method achieves highly competitive novel view synthesis (NVS) quality, recording the top PSNR, SSIM, and LPIPS scores in almost all 'Easy' sequences and outperforming HI-SLAM2 in *upper_turn*. While metrics drop slightly in 'Difficult' sequences, where rapid rotations and bare walls exacerbate monocular depth ambiguity, our overall perceptual fidelity remains strong. Qualitatively (Fig. 8), our method better preserves fine structural details compared to the baselines. In the *xyz* and *sfm* sequences, both MonoGS and HI-SLAM2 suffer from blurred chair boundaries, with MonoGS additionally introducing severe black artifacts on the sofa. In the *circle* sequence, the baselines severely distort the structural geometry beneath the desk, and HI-SLAM2 fails to reconstruct fine wire details. By reflecting natural thermal equilibrium to smooth exclusively non-edge regions, our edge-aware smoothness loss suppresses these artifacts while preserving geometric details. Furthermore, precise camera poses from GE-KLT tracking significantly improve overall rendering fidelity.

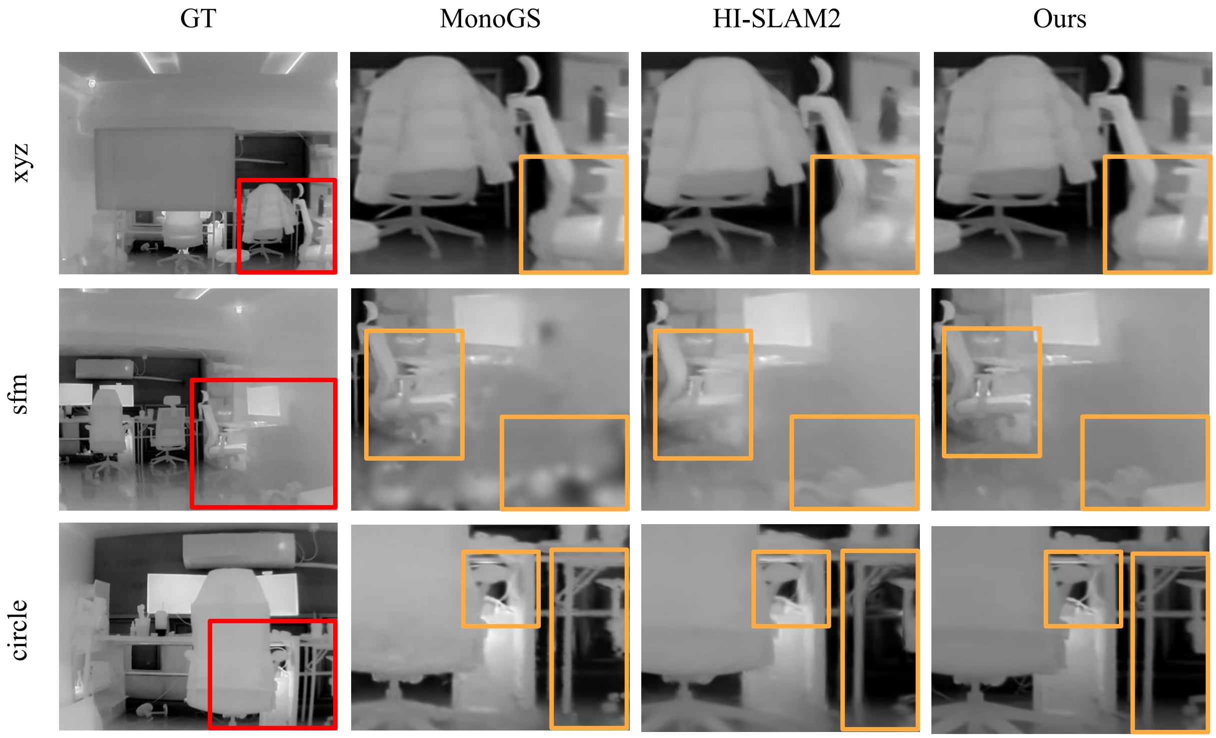


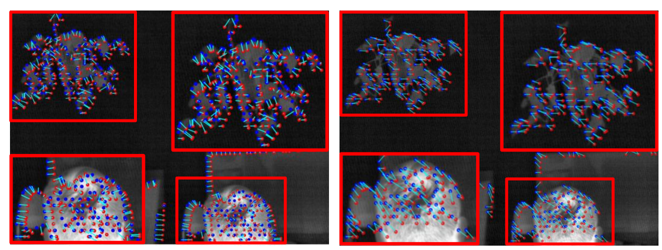
Fig. 8. Qualitative NVS comparison on the custom dataset. From left to right: Ground Truth (GT), MonoGS [1], HI-SLAM2[12], and Ours. (Top) In the *xyz* sequence, MonoGS and HI-SLAM2 blur the boundaries of the chair (orange boxes). (Middle) In the *sfm* sequence, both baselines show blurring around the chair, with MonoGS exhibiting severe black artifacts on the sofa. (Bottom) In the *circle* sequence, HI-SLAM2 loses wire details and both baselines distort the structures under the desk, whereas our method preserves sharp details.

C. Qualitative Analysis of Edge Matching

We provide detailed visual comparisons to validate our core edge matching strategy. For a fair evaluation, the baseline methods, the edge-divergence residual [13] and E-KLT [15], are re-implemented and integrated into our tracking pipeline. We first compare GE-KLT against the edge-divergence residual (Fig. 9). Although the baseline uses spatial proximity and gradient direction consistency, it completely ignores local patch appearances. As highlighted by the red boxes, in regions with dense edge points (e.g., leaf decorations or hair details), the baseline frequently produces inaccurate matches and inconsistent residual directions. Conversely, our two-stage matching leverages local intensity distributions and, combined with rigorous outlier rejection, effectively filters erroneous correspondences and ensures consistent matching directions. Furthermore, Fig. 10 compares GE-KLT with E-KLT [15]. Without geometric constraints, E-KLT allows matched points to drift away from physical boundaries (orange boxes). Our GE-KLT explicitly anchors final matches to detected edge points. This strict structural constraint significantly reduces spurious matches, directly enhancing overall tracking accuracy.

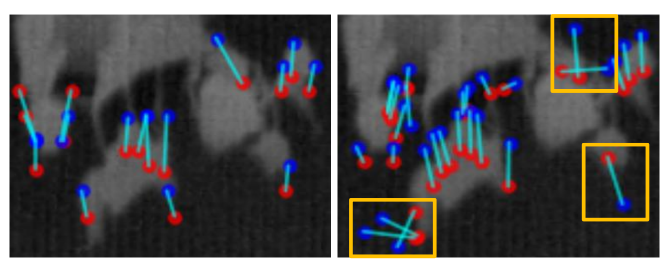
D. Ablation Study

To validate PI-EAT-SLAM's components, we evaluate three challenging multi-spectral sequences featuring varied trajectories (xyz, hfsp), the presence of a person (ps), and aggressive illumination changes (ic) (Table V). First, removing the S-B rescaling module degrades tracking, confirming that physical contrast enhancement is vital for stable thermal edge extraction. Second, omitting GE-KLT tracking causes



(a) Edge-Divergence Residual (b) GE-KLT

Fig. 9. Edge point matching comparison: (a) edge-divergence residual [13] vs. (b) our GE-KLT. Red/blue dots are rendered/GT edge points; cyan lines indicate residual vectors. Compared to the baseline, GE-KLT effectively suppresses erroneous matches and yields highly consistent residual vectors.



(a) GE-KLT (b) E-KLT

Fig. 10. Matching comparison: (a) GE-KLT vs. (b) E-KLT [15]. Orange boxes show E-KLT drifting from actual physical boundaries. In contrast, proposed GE-KLT explicitly anchors matches to valid GT edge points, ensuring accurate boundary alignment.

the most severe degradation (Avg. ATE RMSE drops from 0.0540 to 0.0849 m), highlighting the crucial role of its multi-stage edge matching and dual-constraint outlier rejection in robust pose estimation under extreme noise. Finally, bypassing either the rendering modules (ATF+TCM) or edge-aware smoothing (e-smoothing) decreases accuracy. This confirms that preserving sharp boundaries and mitigating artifacts reinforce geometric consistency, thereby enhancing rendering quality. Utilizing these improved rendered images directly for tracking consequently boosts tracking performance.

TABLE V
QUANTITATIVE ABLATION STUDY ON THE MULTI-SPECTRAL DATASET. RESULTS ARE EVALUATED USING ATE RMSE (M).

Method	ds1-xyz-ps	ds1-hfsp	ds2-xyz-ic	Avg.
w/o S-B rescaling	0.0431	0.0653	0.0861	0.0648
w/o GE-KLT	0.0463	0.0848	0.1236	0.0849
w/o ATF+TCM	0.0426	0.0768	0.0865	0.0686
w/o e-smoothing	0.0436	0.0611	0.0893	0.0647
Ours	**0.0314**	**0.0485**	**0.0840**	**0.0540**

E. Runtime Analysis

We evaluate the operational efficiency of our proposed pipeline by assessing the average computational time on a desktop with a NVIDIA GeForce RTX 3060 GPU. The overall processing time is approximately 1540.37 ms per frame, yielding an operational speed of 0.65 FPS. The runtime breakdown for each core component is as follows: thermal image preprocessing (303.20 ms), GE-KLT tracking (742.42 ms), ATF+TCM (28.52 ms), edge-aware smoothing (98.41 ms), and GS mapping (367.82 ms).

VI. CONCLUSION

We present PI-EAT-SLAM, a novel physics-informed, edge-aware Gaussian Splatting SLAM framework designed for low-contrast, texture-sparse thermal imagery. The proposed method introduces Stefan-Boltzmann law-based thermal rescaling to physically enhance contrast and suppress sensor noise. To overcome the limitations of conventional photometric alignment, we propose GE-KLT tracking. It ensures highly reliable correspondences via a multi-stage edge matching pipeline coupled with rigorous dual-constraint outlier rejection. Furthermore, an edge-aware smoothness loss is applied during mapping to preserve sharp boundaries while enforcing thermal equilibrium. Evaluations on a public multi-spectral dataset and a custom monocular thermal dataset demonstrate superior tracking robustness and competitive novel view synthesis quality against state-of-the-art baselines. Future work includes integrating loop closure for large-scale global consistency and optimizing the pipeline for real-time onboard execution.

REFERENCES

[1] H. Matsuki, R. Murai, P. H. Kelly, and A. J. Davison, “Gaussian splatting slam,” in *IEEE CVPR*, 2024, pp. 18 039–18 048.
[2] N. Keetha, J. Karhade, K. M. Jatavallabhula, G. Yang, S. Scherer, D. Ramanan, and J. Luiten, “Splatam: Splat track & map 3d gaussians for dense rgb-d slam,” in *IEEE CVPR*, 2024, pp. 21 357–21 366.
[3] Y. Wang, H. Chen, Y. Liu, and S. Zhang, “Edge-based monocular thermal-inertial odometry in visually degraded environments,” *IEEE RA-L*, vol. 8, no. 4, pp. 2078–2085, 2023.
[4] M. R. U. Saputra, C. X. Lu, P. P. B. de Gusmao, B. Wang, A. Markham, and N. Trigoni, “Graph-based thermal–inertial slam with probabilistic neural networks,” *IEEE T-RO*, vol. 38, no. 3, pp. 1875–1893, 2021.
[5] Y.-S. Shin and A. Kim, “Sparse depth enhanced direct thermal-infrared slam beyond the visible spectrum,” *IEEE RA-L*, vol. 4, no. 3, pp. 2918–2925, 2019.
[6] Y. Wu, L. Wang, L. Zhang, M. Chen, W. Zhao, D. Zheng, and Y. Cai, “Monocular thermal slam with neural radiance fields for 3d scene reconstruction,” *Neurocomputing*, vol. 617, p. 129041, 2025.
[7] W. Dai, Y. Zhang, S. Chen, D. Sun, and D. Kong, “A multi-spectral dataset for evaluating motion estimation systems,” in *IEEE ICRA*, 2021, pp. 5560–5566.
[8] T. Ye, Q. Wu, J. Deng, G. Liu, L. Liu, S. Xia, L. Pang, W. Yu, and L. Pei, “Thermal-nerf: Neural radiance fields from an infrared camera,” in *IEEE IROS*, 2024, pp. 1046–1053.
[9] H. Gil, M.-H. Jeon, and A. Kim, “Fieldscale: Locality-aware field-based adaptive rescaling for thermal infrared image,” *IEEE RA-L*, vol. 9, no. 7, pp. 6424–6431, 2024.
[10] M. P. Das, L. Matthies, and S. Daftry, “Online photometric calibration of automatic gain thermal infrared cameras,” *IEEE RA-L*, vol. 6, no. 2, pp. 2453–2460, 2021.
[11] X. Chen, W. Dai, J. Jiang, B. He, and Y. Zhang, “Thermal-depth odometry in challenging illumination conditions,” *IEEE RA-L*, vol. 8, no. 7, pp. 3988–3995, 2023.
[12] W. Zhang, Q. Cheng, D. Skuddis, N. Zeller, D. Cremers, and N. Haala, “Hi-slam2: Geometry-aware gaussian slam for fast monocular scene reconstruction,” *IEEE T-RO*, vol. 41, pp. 6478–6493, 2025.
[13] C. Kim, P. Kim, S. Lee, and H. J. Kim, “Edge-based robust rgb-d visual odometry using 2-d edge divergence minimization,” in *IEEE IROS*, 2018, pp. 1–9.
[14] C. Kim, J. Kim, and H. J. Kim, “Edge-based visual odometry with stereo cameras using multiple oriented quadtrees,” in *IEEE IROS*, 2020, pp. 5917–5924.
[15] C. Liu, H. Yu, P. Cheng, W. Sun, J. Civera, and X. Chen, “Pe-vins: Accurate monocular visual-inertial slam with point-edge features,” *IEEE T-IV*, 2024.
[16] J. Shi *et al.*, “Good features to track,” in *IEEE CVPR*, 1994, pp. 593–600.
[17] J. Canny, “A computational approach to edge detection,” *IEEE T-PAMI*, no. 6, pp. 679–698, 2009.
[18] M. Wellons, “The stefan-boltzmann law,” *Physics Department, The College of Wooster, Wooster, Ohio*, vol. 44691, p. 25, 2007.
[19] Q. Chen, S. Shu, and X. Bai, “Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis,” in *ECCV*. Springer, 2024, pp. 253–269.
[20] D. Dhrafani, Y. Liu, A. Jong, U. Shin, Y. He, T. Harp, Y. Hu, J. Oh, and S. Scherer, “Firestereo: Forest infrared stereo dataset for uas depth perception in visually degraded environments,” *IEEE RA-L*, vol. 10, no. 4, pp. 3302–3309, 2025.
[21] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics gems IV*, 1994, pp. 474–485.
[22] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *IEEE ICCV*, 1998, pp. 839–846.
[23] Q. Chen, S. Shu, H. Sun, J. Chen, and X. Bai, “Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis with a large-scale dataset,” *IEEE T-PAMI*, 2026.
[24] R. Lu, H. Chen, Z. Zhu, Y. Qin, M. Lu, L. Zhang, C. Yan, and A. Xue, “Thermalgaussian: Thermal 3d gaussian splatting,” *arXiv preprint arXiv:2409.07200*, 2024.
[25] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, “Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam,” *IEEE T-RO*, vol. 37, no. 6, pp. 1874–1890, 2021.
[26] R. Mur-Artal and J. D. Tardós, “Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras,” *IEEE T-RO*, vol. 33, no. 5, pp. 1255–1262, 2017.