LIV-GaussMap: LiDAR-Inertial-Visual Fusion for Real-time 3D Radiance Field Map Rendering

Sheng Hong[1,*], Junjie He[2,*], Xinhu Zheng[2], Hesheng Wang[4], *Senior Member, IEEE*, Hao Fang[5], *Member, IEEE*, Kangcheng Liu[2†], *Member, IEEE*, Chunran Zheng[3], Shaojie Shen[1], *Senior Member, IEEE*

***Abstract*—We introduce an integrated precise LiDAR, Inertial, and Visual (LIV) multi-modal sensor fused mapping system that builds on the differentiable surface splatting to improve the mapping fidelity, quality, and structural accuracy. Notably, this is also a novel form of tightly coupled map for LiDAR-visual-inertial sensor fusion.**

This system leverages the complementary characteristics of LiDAR and visual data to capture the geometric structures of large-scale 3D scenes and restore their visual surface information with high fidelity. The initial poses for surface Gaussian scenes are obtained using a LiDAR-inertial system with size-adaptive voxels. Then, we optimized and refined the Gaussians by visual-derived photometric gradients to optimize the quality and density of LiDAR measurements.

Our method is compatible with various types of LiDAR, including solid-state and mechanical LiDAR, supporting both repetitive and non-repetitive scanning modes. bolstering structure construction through LiDAR and facilitating real-time generation of photorealistic renderings across diverse LIV datasets. It showcases notable resilience and versatility in generating real-time photorealistic scenes potentially for digital twins and virtual reality, while also holding potential applicability in real-time SLAM and robotics domains.

We release our software and hardware and self-collected datasets on Github[3] to benefit the community.

***Index Terms*—LiDAR, Multi-sensor fusion, Mapping, Radiance Field, 3D Gaussian Splatting.**

I. Introduction

Simultaneous localization and mapping (SLAM), essential for autonomous navigation, combines map construction of unknown environments with tracking an agent's location [2], [3]. Traditional SLAM systems, limited by single sensors like cameras or LiDAR, face challenges like light sensitivity or depth perception issues. Multimodal sensor fusion in SLAM addresses these by integrating data from cameras, LiDAR, and IMUs, enhancing map accuracy and robustness [4]–[6]. Key developments in this area include LiDAR-inertial Visual Odometry (LIVO) and advanced fusion techniques like Kalman Filters, with notable systems like LIC-Fusion, Camvox, LVI-SAM, R2LIVE, R3LIVE, and FAST-LIVO, significantly improving SLAM's efficiency and precision [6]–[11].

However, existing LiDAR-inertial Visual systems are pre-

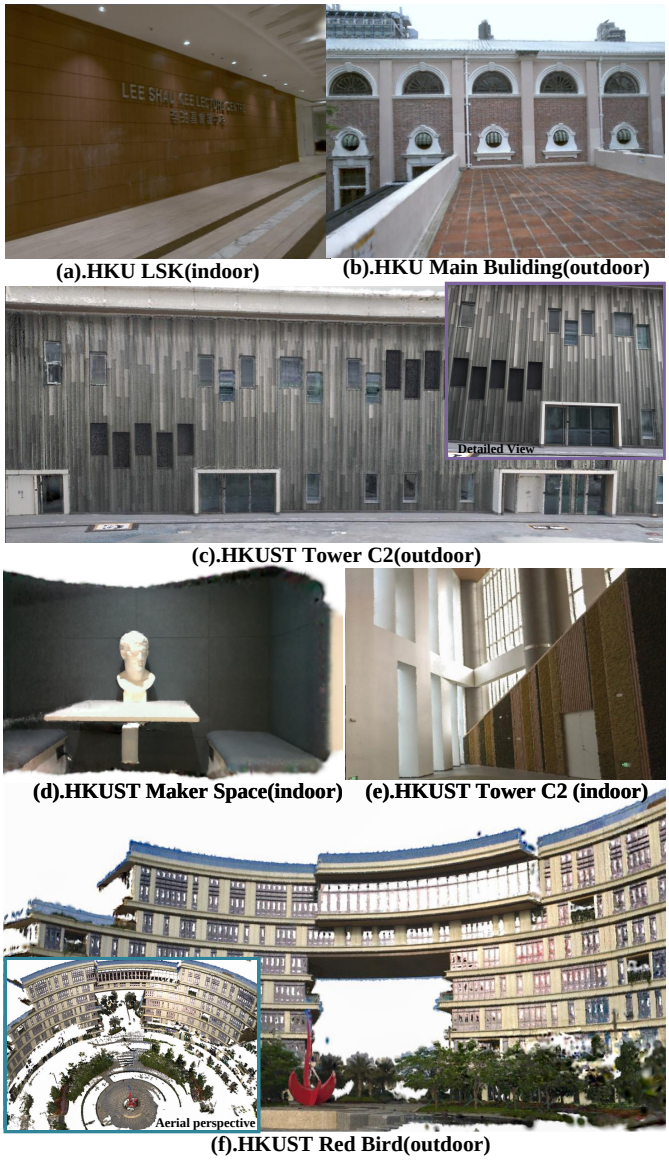

Fig. 1. The real-world experiments were performed in both public datasets and private datasets, including both small-scale indoor environments and large-scale outdoor settings. The image shows our radiance field map of HKU LSK(a), HKU Main Building(b), HKUST GZ Tower C2 outdoor(c) and indoor(e), HKUST GZ Makerspace(d), HKUST GZ Red Bird(f).

* Equal contribution.† Corresponding Author
[1] Department of Electronic Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China.
[2] System hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. The Hong Kong University of Science and Technology, Hong Kong SAR, China.
[3] Department of Mechanical Engineering, The University of Hong Kong, Hong Kong SAR, China.
[4] Department of Automation, Shanghai Jiao Tong University, China.
[5] School of Automation, Beijing Institute of Technology, China.
shongap@connect.ust.hk, kangchengliu@hkust-gz.edu.cn
[3]https://github.com/sheng00125/LIV-GaussMap

(a).The small scale scene (indoor) **(b).The large scale scene (outdoor)**

Fig. 2. The figure shows the aerial perspective of the indoor and outdoor scenes of HKUST GZ Tower C1. In contrast to the vision build-up structure of 3D-GS [1], our approach yields a more refined structure without artifacts.

dominantly designed for scenarios with Lambertian surfaces based on the assumption that the environment exhibits isotropic photometric properties across different viewing directions. Visual information within these systems is typically represented as image patches associated with 3D points [6] or colored pixel [10] [11].

Tracking and mapping in environments with non-Lambertian surfaces like glass or reflective metal are challenging due to their varying reflective properties. Overcoming this requires specialized sensors or algorithms to improve accuracy in such settings.

Recent advancements in novel view synthesis have shown the ability to generate impressive photo-realistic images from new perspectives. These methods employ implicit representations like Neural Radiance Fields (NeRF) [12] or explicit representations such as meshes and signed distance functions, including the emerging technique of 3D Gaussian splatting [1]. By reconstructing the scene's geometric structures while preserving visual integrity with harmonic spherical function, these approaches enable the creation of highly realistic images.

However, in the field of novel view synthesis, the focus on high PSNR often neglects map structure, leading to poor extrapolation performance, crucial for robotics. Techniques like COLMAP and SfM [13] are limited in low-texture scenes. Multimodal sensor fusion improves this, enhancing geometric accuracy and enabling denser, more precise maps.

Overall, the primary contributions of this work can be summarized as follows:

- We propose constructing a dense and precise map structure for the planar surface in the scene by utilizing the Gaussian distribution measurement from the LiDAR-inertial system. This measurement allows us to accurately represent the characteristics of the surface and create a detailed map.
- We propose building up the LiDAR-visual map with differentiable ellipsoidal Gaussians with spherical harmonic coefficients, which implies the visual measurement information from different viewing directions. This approach enables real-time rendering with photorealistic performance, enhancing the accuracy and realism of the map.
- We propose further optimizing the structure of the map by incorporating differentiable ellipsoidal surface Gaussians in order to mitigate the issue of an unreasonable distribution of point clouds caused by the critical inject-angle during scanning, addressing the challenges of unevenly distributed or inaccurately measured point clouds.
- All related software and hardware packages and self-collect datasets will be publicly available to benefit the community.

To our knowledge, this study is the first to utilize multimodal sensor fusion to build a precise and photo-realistic Gaussian map. By combining the accurate map from the LiDAR-inertial system with visual photometric measurements, we achieve a comprehensive and detailed representation of the environment.

Our proposed method has undergone rigorous testing and validation on diverse public real-world datasets, including different types of LiDAR like the mechanical Ouster OS1-128, semi-mechanical Livox Avia, and solid-state Realsense L515. The scene for evaluation covered both indoor (bounded scene) and outdoor (unbounded scene). The experimental results confirm the effectiveness of our algorithm in efficiently capturing and storing image information from multiple viewpoints. This capability enables the rendering of novel views with improved performance.

II. Related works

A. Related work about Mapping with Multi-modal Sensor

In the realm of robotics, multi-modal sensor fusion for localization, such as LiDAR-inertial visual odometry(LIVO), is being extensively researched. LiDAR can deliver accurate geometric measurements of real-world environments, while cameras provide detailed 2D imagery of textures and appearances of the environment. Meanwhile, inertial navigation systems supply high-frequency motion measurements. Integrating these sensors is considered ideal for robotic applications. Numerous works in this area, such as LIC-Fusion [7], R2LIVE [10], LVI-SAM [9], Camvox [8], R3LIVE [11] and FAST-LIVO [6] have been contributing significantly by enhancing perception capabilities in robotics.

Among them, R3LIVE [11] and FAST-LIVO [6] adopt the tightly coupled iterative error state Kalman filtering methods for multi-modal sensor fusion. They provide accurate and real-time odometry and generate colored point clouds.

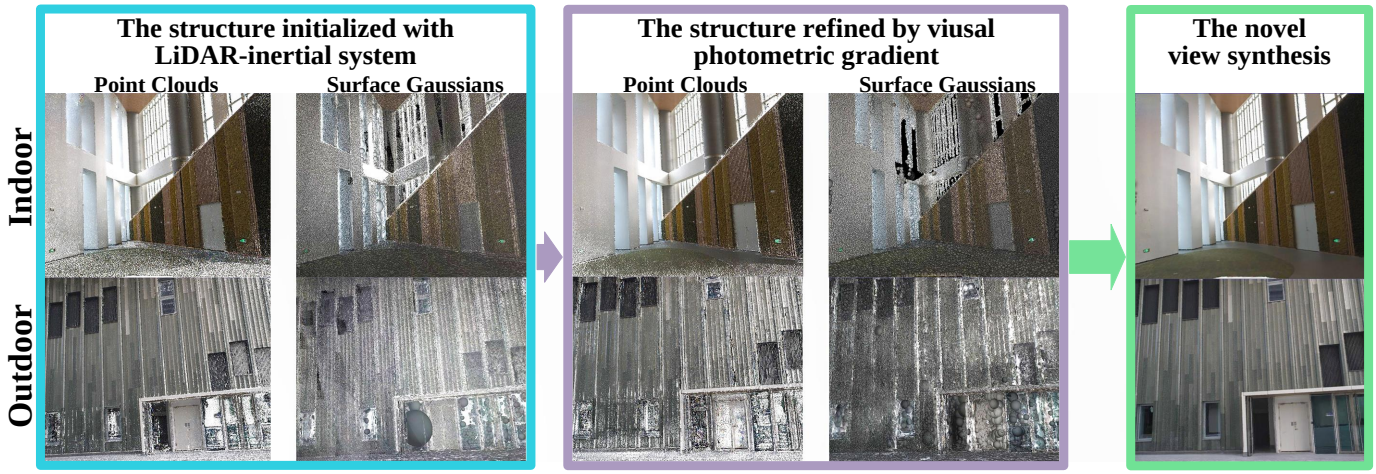

Fig. 3. The construction process of the map is illustrated in the above figure. (1). Initially, the Gaussians of the scene are derived from a Kalman-filtered LiDAR-inertial system. The surfaces of 3D objects within the scene are estimated using LiDAR measurements. The Gaussians expand along the surface, resulting in an initial colored point cloud. This further develops into the ellipsoidal surface Gaussians. (2). We then enhance the Gaussian distribution utilizing photometric gradients, leading to an optimized point cloud and an optimized Surface Gaussian. This optimized map allows us to synthesize new views with precise photometry and generate maps that are devoid of any gaps.

R3LIVE uses photometric error in RGB-colored point clouds, while FAST-LIVO utilizes warped image patches from diverse viewpoints, similar to SVO. These approaches highlight the flexibility of multi-modal sensor fusion in robotics. LIV-based SLAM systems can generate high-density colored point clouds for realistic visualization. However, these point clouds may have holes and lack photometric realism. Moreover, for anisotropic, non-Lambertian surfaces like glass and metal, the RGB values can vary across different viewpoints, leading to a blurred RGB point cloud map. In [14], an efficient LIO method is presented, employing adaptive voxels with plane features for improved scene mapping and precise LiDAR scan registration.

In the context of voxel-based mapping and odometry methods, [11] and [6] map the world with fixed-size voxels, while [14] build up the voxel with adaptive size. These approaches model the surface in a scene with Gaussian distribution, which resembles surface splatting [1] [15] [16] used for novel view synthesis in computer graphics and 3D visualization.

B. Related Work about Novel View Synthesis

Novel view synthesis has progressed with continuous radiance field modeling, using explicit representations like meshes, point clouds, SDFs, or implicit ones like NeRF [12]. These methods, differing from traditional SLAM systems with discrete point clouds, create more photorealistic images by treating scenes as continuous, viewpoint-dependent functions. Instant-NGP [17] further accelerates this with its multiresolution grid structure, enabling real-time rendering and faster training. Mip-NeRF 360 [18] addresses unbounded scenes and sampling issues with nonlinear parametrization and new regularizers. Despite implicit representations using neural networks for high-fidelity, low-memory synthesis, they remain computationally intensive. Recent efforts focus on explicit map representations, like using spherical harmonics for voxel-based volumetric density.

PlenOctrees, introduced by Yu et al. [19], utilize volumetric rendering with spherical harmonics (SHs) to model rays from various directions, offering a compact and efficient way to represent complex 3D scenes. Plenoxels, proposed by Fridovich-Keil et al. [20], represent scenes as sparse 3D grids using SHs, optimized through gradient methods without the need for neural networks, significantly reducing computational requirements and achieving real-time rendering speeds 100 times faster than NeRF. The concept of splatting-based rendering, originating from Zwicker et al.'s surface splatting [16], has evolved through differentiable surface splatting for point-based geometry by Wang et al. [21], and further advancements in optimizing gradients for SH coefficients in splattings by Zhang et al. [15]. Most recently, Kerbl et al. [1] have developed a method to simulate spatial object surfaces as anisotropic Gaussian-distributed splattings, enabling image synthesis from novel viewpoints.

III. Methodology

Our system, illustrated in Fig. 4, integrates hardware and software components. Hardware-wise, it features a hardware-synchronized LiDAR-inertial sensor paired with a camera, ensuring precise synchronization of LiDAR point clouds and image captures for accurate data alignment and fusion.

Software-wise, the process starts with LiDAR-inertial odometry [14] for localization, using a size-adaptive voxel map to represent planar surfaces. LiDAR point clouds are segmented into voxels, where the covariance of planes is computed for initial elliptical splatting estimates (see Fig. 3). The final step involves refining spherical harmonic coefficients and LiDAR Gaussian structures using images from various perspectives, leveraging photometric gradients. This approach produces a photometrically accurate LiDAR-visual map, enhancing mapping precision and visual realism.

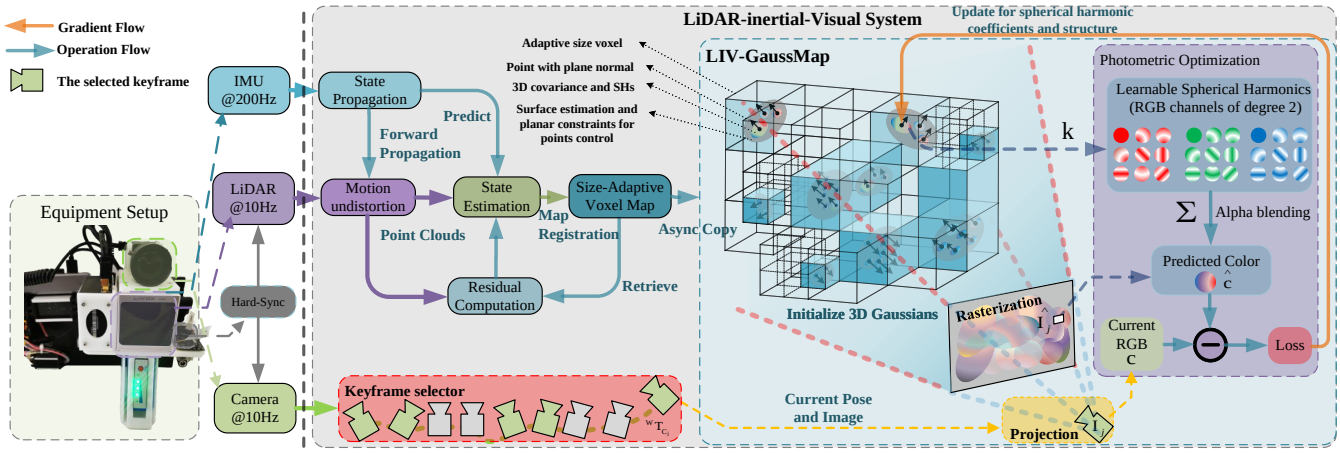

Fig. 4. The left of the image illustrates the sensory input and our equipment setup. It features an external sensor assembly comprising a synchronized LiDAR-Inertial sensor (Livox Avia) paired with a camera. On the right side, we present our algorithmic pipeline, which includes: 1. The initial representation of the scene is derived from a IESKF-based LiDAR-inertial system with size adaptive voxel, providing an initial Gaussian structure for the scene. 2. Subsequently, we optimize the Gaussians structure and spherical harmonic coefficient by photometric gradients. This involves calculating rasterization loss using images to refine the scene representation further.

A. Initialization of Gaussians with LiDAR Measurement

Initially, we employ size-adaptive voxels to partition the LiDAR point cloud, drawing inspiration from the Octree approach discussed in [14]. We meticulously construct a nuanced Gaussian surface. Initially, we apply the size-adaptive voxels to partition the LiDAR point cloud akin to Octree inspired by [14]. Our adaptiveness of voxel partitioning is determined based on evaluating a certain parameter η, which serves as an indicator to judge if a voxel has a surface with planar characteristics inside. To obtain a more precise map with a normal vector of Gaussian surface, we allow for smaller voxels and further subdivision into finer levels. If the voxel is divided small enough through multiple subdivisions, even curved surfaces can be approximated.

The voxel can be characterized by its average position $\overline{\mathbf{p}}$, the normal vector $\mathbf{n}$, and the covariance matrix $\mathbf{\Sigma}_{\mathbf{n},\overline{\mathbf{p}}}$ inside the voxel.

$$\overline{\mathbf{p}} = \frac{1}{N}\sum_{i=1}^{N} {}^{w}\mathbf{p}_i \tag{1}$$

The covariance of voxel $\mathbf{\Sigma}_{\mathbf{n},\overline{\mathbf{p}}}$ can be calculated as below, which indicates the distribution of the points ${}^{w}\mathbf{p}_i$:

$$\mathbf{\Sigma}_{\mathbf{n},\overline{\mathbf{p}}} = \frac{1}{N}\sum_{i=1}^{N} ({}^{w}\mathbf{p}_i - \overline{\mathbf{p}})\,({}^{w}\mathbf{p}_i - \overline{\mathbf{p}})^T \tag{2}$$

We denote the eigenvector $\mathbf{n}$, which is regarded as the normal vector of the planar surface, for the covariance $\mathbf{\Sigma}_{\mathbf{n},\overline{\mathbf{p}}}$ of this hypothetical Gaussian plane [14]. The corresponding eigenvalues λ represent the distribution of this Gaussian plane in each direction. If the η, which indicates the thickness of the planar surface, is still significant, further subdivision is performed.

$$\eta = \frac{\lambda_{min}}{\sqrt{\lambda_{\mathbf{mid}}{}^2 + \lambda_{min}^2 + \lambda_{\mathbf{max}}{}^2}} \tag{3}$$

The distribution matrix $\mathbf{\Sigma}_{\mathbf{n},\overline{\mathbf{p}}}$ is calculated to determine the approximate shape and pose of the point cloud, which contains the pose of the surface Gaussian. However, to seamlessly integrate these LiDAR points with surrounding points and ensure hole-free proportional scaling that upholds the integrity of the original data. We introduce a scaling factor α_i for each point, which is determined by the point density. This scaling factor allows for the rescaling of the points accordingly.

$$\mathbf{\Sigma}_{\mathbf{w}_{\mathbf{p}_i}} = \alpha_i \mathbf{\Sigma}_{\mathbf{n},\overline{\mathbf{p}}} \tag{4}$$

We define the 3D radiance field of the LiDAR point cloud with the elegant form of an ellipsoidal Gaussian, represented by the following equation:

$$G(x) = e^{-\frac{1}{2}({}^{w}\mathbf{p}_i - x)^T \mathbf{\Sigma}_{\mathbf{w}_{\mathbf{p}_i}}{}^{-1} ({}^{w}\mathbf{p}_i - x)} \tag{5}$$

B. Spherical Harmonic Coefficient Optimization and Map Structure Refinement with Photometric Gradients

The structure provided by the LiDAR-inertial system is further refined with visual photometric gradients for enhanced mapping. Further, we utilize high-order spherical harmonics, akin to those in computer graphics, for depicting view-dependent radiance surfaces.

We utilize second-degree spherical harmonics (SHs) [22], which requires a total of 27 harmonic coefficients for each Gaussian. And it is a balance between complexity (and thus computational cost) and accuracy. A more photorealistic map can be obtained by optimizing the spherical harmonic coefficients of LiDAR Gaussian through a photometric gradient. This refined map enables real-time rendering with improved interpolation and extrapolation for photorealistic mapping.

The point in the world frame is ${}^{w}\mathbf{p}_i$, and the pose of the LIV system is ${}^{w}\mathbf{T}_{C_n}$. The viewing direction of point ${}^{w}\mathbf{p}_i$ from the pose ${}^{w}\mathbf{T}_{C_n}$ can be calculated as:

$$ {}^{C_n}\mathbf{v}_i = \frac{{}^{w}\mathbf{T}_{C_n}^{-1} \cdot {}^{w}\mathbf{p}_i}{\|{}^{w}\mathbf{T}_{C_n}^{-1} \cdot {}^{w}\mathbf{p}_i\|} \tag{6} $$

$$ \theta = \arccos\left(\frac{{}^{C_n}\mathbf{v}_{iz}}{\sqrt{{}^{C_n}\mathbf{v}_{ix}^2 + {}^{C_n}\mathbf{v}_{iy}^2 + {}^{C_n}\mathbf{v}_{iz}^2}}\right) \tag{7} $$

$$ \phi = \arctan 2({}^{C_n}\mathbf{v}_{iy}, {}^{C_n}\mathbf{v}_{ix}) \tag{8} $$

The spherical harmonics function is sensitive to the viewing direction.

$$ c(\theta,\phi) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} k_\ell^m \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta)e^{im\phi} \tag{9} $$

where $P_\ell^m(\cos\theta)e^{im\phi}$ represents the Legendre polynomials.

As the Fig.3 shows, this LIV system initially generates a dense point cloud populated with 3D Gaussians, each characterized by a position ${}^{w}\mathbf{p}_i$, a covariance matrix $\mathbf{\Sigma}_{w_{\mathbf{p}_i}}$, a normal vector n, and a color $\mathbf{c}$ and a opacity α.

Consequently, The projection of a LiDAR point cloud from the world frame to the camera frame C_n at image plane ${}^{C_n}\boldsymbol{q}_i$ can be written as:

$$ {}^{C_n}\mathbf{q}_i = \boldsymbol{\pi}({}^{w}\mathbf{T}_{C_n}^{-1} \cdot {}^{w}\mathbf{p}_i) \tag{10} $$

To train a point cloud model that predicts the image I_n. We employ a loss function consisting of an MSE to optimize the structure and the spherical harmonic coefficients of point clouds, that is

$$ \mathcal{L} = (1-\lambda)\sum_{n=1}^{N}\sum_{q\in\mathcal{R}}\left\|I_n(q) - \hat{I}_n(q)\right\| + \lambda\mathcal{L}_{\textbf{D-SSIM}} \tag{11} $$

C. Adaptive Control of 3D Gaussian map

The structure derived from the LiDAR-inertial system is not flawless. It may encounter difficulties in measuring surfaces made of glass or areas that have been either excessively or insufficiently scanned.

To tackle these concerns, we employ structure refinement to address under-reconstruction and over-dense scenarios.

In situations where geometric features are not yet well reconstructed (under-reconstruction), noticeable positional gradients can arise within the view-space. In our experiments, we establish a predefined threshold value to identify regions that require densification. We replicate the neighboring Gaussians and then employ photometric gradient optimization to precisely position these replicated Gaussians for completion.

For over-dense area, we regularly evaluate its net contribution, as shown in Eq. (12), and eliminate excessively nonessential regions. This effectively reduces redundant points in the map and enhances optimization efficiency.

The net contribution $A_i^j(u)$ of each point is determined by:

$$ A_i^j(u) = \alpha_i^j(u)\prod_{k=1}^{i-1}(1-\alpha_k^j(u)) \tag{12} $$

Where, the opacity $\alpha_i^j(u)$ for each point is calculated using the formula:

$$ \alpha_i^j(u) = \frac{1}{\sqrt{2\pi r^2}} e^{-\frac{\|p_i^j - u\|^2}{2r^2}} \tag{13} $$

For efficiency, we consider only points within a specific maximum distance from u and with significant opacity.

D. Novel View Synthesis with Gaussians

By employing rasterization [16] for the synthesis of images from a Gaussian cloud generated by LiDAR, the novel view of the image $\hat{I}_j(u)$ can be synthesized through alpha blending using the following equation:

$$ \hat{I}_n(u) = \sum_{i=1}^{M} A_i^j(u)c_i^j \tag{14} $$

IV. Real-world experiments

As shown in Table I, the detailed device configurations for the four evaluation datasets are presented. To thoroughly evaluate the effectiveness of our algorithm, we purposely conducted tests on two publicly accessible datasets and two proprietary datasets that encompass a wide range of LiDAR modalities. Specifically, we leveraged the FusionPortable dataset [23], which features repetitive scanning LiDAR, and the FAST-LIVO dataset [6], which includes non-repetitive LiDAR data from public sources. In comparison to existing datasets, ours offers a comprehensive array of LiDAR modalities captured in both indoor and outdoor environments, ensuring robust hardware synchronization and accurate intrinsic [24] and extrinsic [25] parameter calibration. Additionally, we provide ground truth structures in the form of point clouds to facilitate structure accuracy evaluation. For the execution of our Mapping system, we employ a high-performance desktop computer powered by an Intel Core i9 12900K 3.50GHz processor and a single NVIDIA GeForce RTX 4090.

The experiment results highlighted a well-optimized map structure, surpassing the original novel view synthesis method. With the photometric optimization for the splattings, our proposed algorithm exhibits competitive performance in terms of both interpolated and extrapolated PSNR. The fact that our LIV-based rendering system outperforms the original novel view synthesis system in terms of structure further underscores the effectiveness of our approach, as shown in Fig. 2.

In the subsequent section, we conducted comparative and ablation experiments to evaluate the optimization factors, which revealed a considerable improvement in both PSNR and the structural score.

A. Evaluation for Novel View Synthesis with previous work

As shown in Fig. 6, we evaluated the performance of our mapping system on a real-world dataset for rendering quality against other state-of-the-art frameworks, namely Plenoxel [20], F2-NeRF [3], and 3d Gaussian splatting [1]. As shown in Table II and the boxplot shown in Fig. 5, our framework achieved a significant improvement of 5dB in PSNR for extrapolation, indicating superior rendering quality. Additionally, we observed a competitive performance in interpolation PSNR as well.

Our algorithm demonstrates a comparable level of Peak Signal-to-Noise Ratio (PSNR) with 3d Gaussian splatting among other state-of-the-art (SOTA) algorithms. Besides, we achieve a remarkable score for extrapolation. This can be attributed to the inherent advantage of our LiDAR system, which offers a relatively precise structural observation.

TABLE I
SPECIFICATIONS OF LIDAR-INERTIAL-VISUAL SYSTEM IN TESTED DATASETS

Dataset		FAST-LIVO [6]	FusionPortable [23]	Our Device I	Our Device II
LiDAR	Device name	Livox Avia	Ouster OS1-128	RealSense L515	Livox Avia
	Points per second	240,000	2,621,440	23,000,000	240,000
	Scanning mechanism	Mechanical, non-repetitive	Mechanical, repetitive	Solid-state	Mechanical, non-repetitive
	Range	$3\ \text{m} - 450\ \text{m}$	$1\ \text{m} - 120\ \text{m}$	$9\ \text{m}$ - $25\ \text{m}$	$3\ \text{m} - 450\ \text{m}$
	Field of View	$70.4^\circ \times 77.2^\circ$	$45^\circ \times 360^\circ$	$70^\circ \times 55^\circ$	$70.4^\circ \times 77.2^\circ$
	IMU	BM1088	ICM20948	BMI085	BMI088
Camera	Device name	MV-CA013-21UC	FILR BFS-U3-31S4C	RealSense L515	MV-CA013-21UC
	Shutter mode	Global shutter	Global shutter	Rolling shutter	Global shutter
	Resolution	1280×1024	1024×768	1920×1080	1280×1024
	Field of View	$72^\circ \times 60^\circ$	$66.5^\circ \times 82.9^\circ$	$70^\circ \times 43^\circ$	$72^\circ \times 60^\circ$
Synchronization		✓	✓	✓	✓
Grouth Truth of Structure		×	✓	✓	✓
Dataset Sequence		HKU_LSK(indoor) HKU_MB(outdoor)	HKUST_indoor	UST_RBMS	UST_C2_indoor UST_C2_outdoor

TABLE II
QUANTITATIVE EVALUATION OF OUR METHOD WITH PREVIOUS WORK

	PSNR[dB]↑ (Interpolate)	SSIM↑ (Interpolate)	LPIPS↓ (Interpolate)	PSNR↑ (Extrapolate)	SSIM↑ (Extrapolate)	LPIPS↓ (Extrapolate)
Plenoxel [20]	26.744	0.844	0.452	12.916	0.628	0.575
Mip-NeRF-360 [18]	28.446	0.820	0.444	19.213	0.726	0.526
F2-NeRF [26]	32.556	0.941	0.193	19.100	0.764	0.387
3d Gaussian Splatting [1]	31.900	0.913	0.241	15.112	0.647	0.503
Our method	32.787	0.926	0.190	19.220	0.803	0.331

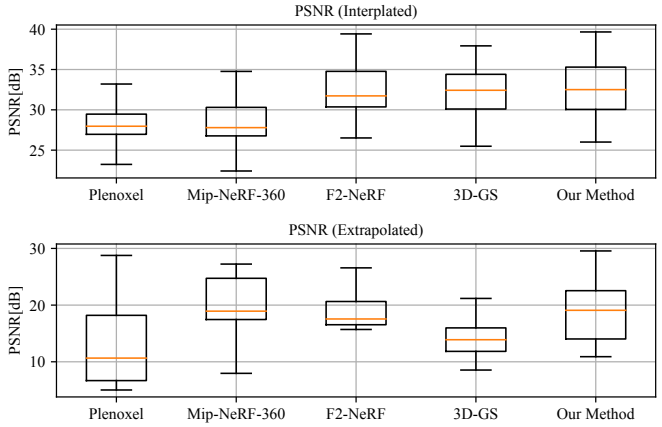

Fig. 5. This box plot illustrates the comparative performance of the leading method and our approach in terms of interpolation and extrapolation across datasets by PSNR values.

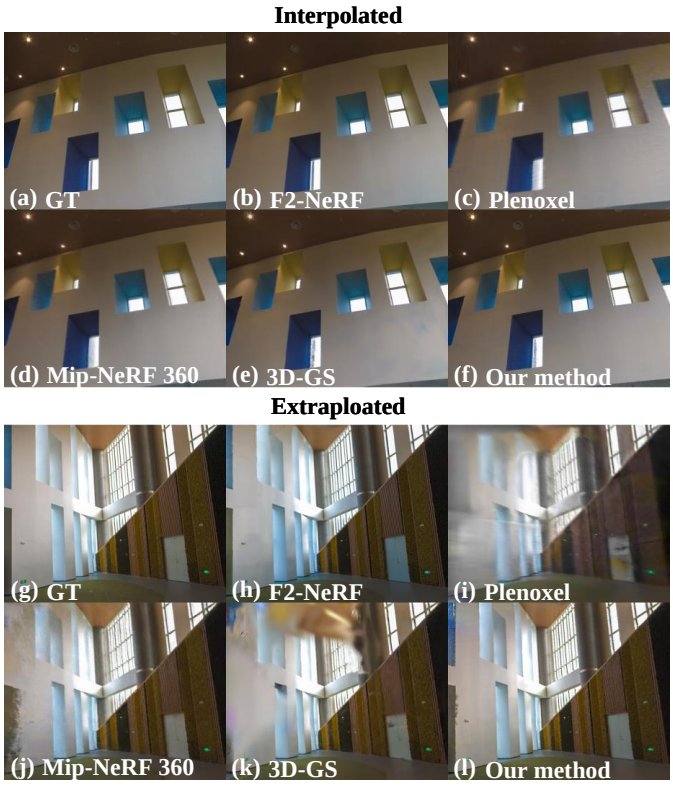

Fig. 6. We present a comprehensive comparison between our proposed method and the state-of-the-art technique, showcasing the results of both interpolation and extrapolation for synthesizing novel viewpoints. The upper row exhibits interpolated views, while the bottom row demonstrates extrapolated viewpoint synthesis.

B. Ablation Study for rendering performance with LiDAR structure

Based on the experiments above, we have discovered that 3D Gaussian splatting exhibits remarkable competitiveness. Consequently, our subsequent experiments will primarily revolve around comparing the 3D Gaussian splatting technique with our LiDAR-assisted Gaussian structure construction approach. As shown in Table III, to validate the effectiveness of our algorithm, we progressively integrated our optimized methods and monitored the corresponding changes in PSNR. We designed several comparative experiments as the following cases for ablation analysis.

Case I: Implemented 3D-GS as the baseline. Case II: LiDAR initialization of Gaussians, showing map reconstruction

TABLE III
ABLATION STUDY FOR MAP STRUCTURE OPTIMIZATION

Metric	Method	HKU_MB	HKU_LSK	UST_C2_outdoor	UST_C2_indoor	UST_RBMS	Avg.
PSNR[db]↑ (Interpolated)	Case I	24.390	31.222	31.843	31.721	31.663	30.168
	Case II	24.341	25.964	31.983	29.625	30.211	28.425
	Case III	24.240	31.045	33.229	31.975	31.047	30.307
	Case IV	25.140	31.597	33.644	32.726	31.277	30.877
SSIM↑ (Interpolated)	Case I	0.793	0.798	0.897	0.916	0.872	0.856
	Case II	0.814	0.780	0.895	0.891	0.864	0.849
	Case III	0.809	0.804	0.909	0.918	0.868	0.862
	Case IV	0.825	0.805	0.916	0.926	0.870	0.868
LPIPS↓ (Interpolated)	Case I	0.316	0.277	0.115	0.219	0.338	0.253
	Case II	0.304	0.292	0.147	0.219	0.358	0.264
	Case III	0.301	0.273	0.101	0.195	0.349	0.244
	Case IV	0.296	0.259	0.094	0.190	0.341	0.236
PSNR[db]↑ (Extrapolate)	Case I	15.144	23.831	24.426	18.657	23.868	21.185
	Case II	16.503	22.400	25.653	20.511	23.792	21.772
	Case III	16.178	24.821	25.047	18.964	24.879	21.978
	Case IV	16.530	24.808	25.912	19.220	25.545	22.403
SSIM↑ (Extrapolate)	Case I	0.403	0.680	0.570	0.766	0.849	0.654
	Case II	0.441	0.657	0.674	0.771	0.847	0.679
	Case III	0.451	0.686	0.612	0.775	0.848	0.675
	Case IV	0.470	0.684	0.648	0.801	0.851	0.691
LPIPS↓ (Extrapolate)	Case I	0.530	0.402	0.307	0.370	0.389	0.399
	Case II	0.494	0.355	0.302	0.314	0.393	0.372
	Case III	0.482	0.356	0.314	0.341	0.382	0.375
	Case IV	0.479	0.348	0.275	0.336	0.371	0.362
Cost time[min]	Case I	26m9s	16m58s	24m16s	14m15s	20m43s	20m28s
	Case II	34m14s	17m3s	25m55s	17m38s	24m5s	23m47s
	Case III	19m37s	14m19s	16m26s	12m20s	18m54s	16m19s
	Case IV	18m19s	13m41s	16m33s	13m4s	18m46s	16m5s

without visual optimizations. Case III: Enhanced Case II with photometric gradients to optimize point cloud distribution, improving map accuracy and robustness. Case IV: Further refinement using photometric gradients for Gaussian pose optimization, enhancing map quality.

Comparing Case I and II, LiDAR performance varies with scene complexity. In complex structures like "HKU_MB", LiDAR's accuracy decreases, potentially lowering PSNR. In simpler scenes like "UST_C2_outdoor", LiDAR achieves precise estimations, enhancing PSNR, especially in extrapolation tasks. In Case III, optimizing point cloud distribution speeds up the process but may reduce PSNR.

Ultimately, our method (Case IV) enhances PSNR by refining the map structure, consistently outperforming 3D-GS across all scenes in both interpolation and extrapolation. We also evaluated our method with solid-state LiDAR (RealSense L515). Given its restricted measurement range, we only conducted experiments in indoor scenes. The results demonstrate that our approach consistently preserves a superior level of PSNR.

C. Structure Reconstruction Evaluation

Our study presented both qualitative and quantitative results, highlighting the effectiveness of using LiDAR for initial structure optimization (Fig 2). Quantitatively, we evaluated our approach using Chamfer Discrepancy (CD), Earth Mover Distance (EMD), and F-score (Tab IV), finding significant improvements in these metrics with LiDAR-based initialization. While the use of photometric optimization Gaussian distribution slightly reduced structural quality, the introduction of Gaussian pose refinement showed mixed results: it improved CD (Chamfer Discrepancy) [27] and EMD(Earth Mover Distance) [28] but negatively impacted the F-score [29]. Despite some trade-offs in structural integrity for better PSNR, our method overall demonstrated superior structural metrics compared to purely visual approaches.

TABLE IV
ABLATION STUDY OF DIFFERENT DESIGN CHOICES ON FUSIONPORTABLE DATASET [23]

	CD [27]↓	EMD [28]↓	F-score [29]↑
Case I	0.149	0.698	0.544
Case II	0.114	0.553	0.807
Case III	0.109	0.614	0.682
Case IV	0.107	0.435	0.751

V. Conclusion

We propose LiDAR-Inertial-Visual fused real-time 3D radiance field mapping system that capitalizes on the fusion of LiDAR-inertial visual multi-modal sensors.

Our approach leverages the precise surface measurement capabilities of LiDAR, coupled with the adaptive voxel feature innate to the LiDAR-inertial system, facilitating rapid initial scene structure acquisition. However, it's inevitable to encounter critical injection angles during LiDAR scanning, which can lead to unreasonably distributed or inaccurately measured point clouds. To address this, we utilize the photometric gradient derived from visual observations to further optimize the LiDAR structure, thereby enhancing PSNR performance. Additionally, our proposed LiDAR-Visual map seamlessly integrates LiDAR measurements and visual observations from all viewing directions, offering real-time rendering capabilities.

Through extensive real-world experiments, we have consistently demonstrated that our algorithm achieves superior geometric structures. Moreover, it produces novel view images with higher PSNR than other state-of-the-art visual-based methods, both for extrapolated and interpolated poses.

References

[1] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics (ToG)*, vol. 42, no. 4, pp. 1–14, 2023.

[2] K. Liu and M. Cao, "Dlc-slam: A robust lidar-slam system with learning-based denoising and loop closure," *IEEE/ASME Transactions on Mechatronics*, 2023.

[3] Q. Xu, Z. Xu, J. Philip, S. Bi, Z. Shu, K. Sunkavalli, and U. Neumann, "Point-nerf: Point-based neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5438–5448.

[4] K. Liu and H. Ou, "A light-weight lidar-inertial slam system with high efficiency and loop closure detection capacity," in *2022 International conference on advanced robotics and mechatronics (ICARM)*. IEEE, 2022, pp. 284–289.

[5] K. Liu, "A robust and efficient lidar-inertial-visual fused simultaneous localization and mapping system with loop closure," in *2022 12th international conference on CYBER technology in automation, control, and intelligent systems (CYBER)*. IEEE, 2022, pp. 1182–1187.

[6] C. Zheng, Q. Zhu, W. Xu, X. Liu, Q. Guo, and F. Zhang, "Fast-livo: Fast and tightly-coupled sparse-direct lidar-inertial-visual odometry," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 4003–4009.

[7] X. Zuo, P. Geneva, W. Lee, Y. Liu, and G. Huang, "Lic-fusion: Lidar-inertial-camera odometry," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5848–5854.

[8] Y. Zhu, C. Zheng, C. Yuan, X. Huang, and X. Hong, "Camvox: A low-cost and accurate lidar-assisted visual slam system," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5049–5055.

[9] T. Shan, B. Englot, C. Ratti, and D. Rus, "Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 5692–5698.

[10] J. Lin, C. Zheng, W. Xu, and F. Zhang, "R2 live: A robust, real-time, lidar-inertial-visual tightly-coupled state estimator and mapping," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7469–7476, 2021.

[11] 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)*. IEEE, 2022, pp. 10 672–10 678.

[12] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[13] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.

[14] C. Yuan, W. Xu, X. Liu, X. Hong, and F. Zhang, "Efficient and probabilistic adaptive voxel mapping for accurate online lidar odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8518–8525, 2022.

[15] Q. Zhang, S.-H. Baek, S. Rusinkiewicz, and F. Heide, "Differentiable point-based radiance fields for efficient view synthesis," in *SIGGRAPH Asia 2022 Conference Papers*, 2022, pp. 1–12.

[16] M. Zwicker, H. Pfister, J. Van Baar, and M. Gross, "Surface splatting," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 371–378.

[17] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics (ToG)*, vol. 41, no. 4, pp. 1–15, 2022.

[18] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5470–5479.

[19] A. Yu, R. Li, M. Tancik, H. Li, R. Ng, and A. Kanazawa, "Plenoctrees for real-time rendering of neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5752–5761.

[20] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa, "Plenoxels: Radiance fields without neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5501–5510.

[21] W. Yifan, F. Serena, S. Wu, C. Öztireli, and O. Sorkine-Hornung, "Differentiable surface splatting for point-based geometry processing," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, pp. 1–14, 2019.

[22] B. Cabral, N. Max, and R. Springmeyer, "Bidirectional reflection functions from surface bump maps," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, 1987, pp. 273–281.

[23] J. Jiao, H. Wei, T. Hu, X. Hu, Y. Zhu, Z. He, J. Wu, J. Yu, X. Xie, H. Huang, R. Geng, L. Wang, and M. Liu, "Fusionportable: A multi-sensor campus-scene dataset for evaluation of localization and mapping accuracy on diverse platforms," 2022.

[24] 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*. IEEE, 2013, pp. 1280–1286.

[25] C. Yuan, X. Liu, X. Hong, and F. Zhang, "Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7517–7524, 2021.

[26] P. Wang, Y. Liu, Z. Chen, L. Liu, Z. Liu, T. Komura, C. Theobalt, and W. Wang, "F2-nerf: Fast neural radiance field training with free camera trajectories," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4150–4159.

[27] T. Nguyen, Q.-H. Pham, T. Le, T. Pham, N. Ho, and B.-S. Hua, "Point-set distances for learning representations of 3d point clouds," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 478–10 487.

[28] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 605–613.

[29] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation," in *Australasian joint conference on artificial intelligence*. Springer, 2006, pp. 1015–1021.