Thermal-Depth Odometry in Challenging Illumination Conditions

Xingxin Chen, Weichen Dai, Jiajun Jiang, Bin He, and Yu Zhang

***Abstract*—Traditional Visual Odometry (VO) methods that utilize visible cameras frequently degrade in challenging illumination environments. Alternative vision sensors such as thermal cameras are promising for all-day navigation since the delivered thermal images are invariant to ambient illumination. However, traditional VO techniques cannot be directly translated to the thermal domain due to poor thermal image quality. Besides, the thermal cameras stop image capture during the unique imaging mechanism (e.g., Non-Uniformity Correction (NUC)), making the thermal VO easily lose tracking. In this letter, we propose a thermal-depth odometry method that can fuse information from both types of sensors, thermal and depth cameras. The system front-end estimates 6-DoF camera motion via a semi-direct framework, fully exploiting thermographic data cues from raw thermal images. The depth information is aligned with the thermal images by extrinsic parameters to enhance the robustness of motion estimation. To overcome the challenge from the NUC, the proposed method introduces an NUC handling module, which can conduct pose estimation by registering multiple point clouds generated from depth images. The proposed method is evaluated on public datasets. The results demonstrate that the proposed method can provide competitive localization performance under different illumination.**

***Index Terms*—Localization, SLAM.**

I. Introduction

ROBUST and accurate localization is an essential capability for autonomous systems. In GPS-denied environments, visual odometry has become a popular navigation solution as the low power consumption and portability of cameras. However, traditional VO methods based on visible cameras lack robustness in environments with challenging illumination, such as underground tunnels and dark rooms. In contrast, long-wave infrared thermal cameras can capture thermal infrared radiation information, which is irrelevant to ambient lighting conditions. Therefore, visual odometry based on thermal cameras gained much attention from the communities.

Some efforts have been made toward thermal odometry. However, robust and accurate ego-motion estimation remains challenging for thermal imaging systems. One hindrance is the Non-Uniformity Correction (NUC). NUC is required for noise reduction and periodically suspends the camera operation for about half to one second, depending on the specific camera and the complexity of the correction algorithm. The NUC process can be triggered automatically by the camera's software based on time intervals and temperature changes, or manually upon command. The interruption of image data causes a significant viewpoint change and makes the odometry prone to lose tracking. To address this issue, thermal cameras are typically complemented by an Inertial Measurement Unit (IMU) [1], [2]. However, IMU propagation faces the challenge of accumulated error due to the inherent noise and bias in the IMU measurements [3], [4]. Another promising alternative is depth cameras. During the NUC operation, the camera pose can be estimated by the point clouds generated from the depth images. Incorporating a depth camera with a thermal camera has several advantages. (1) Depth measurements provide metric scale and a simplified initialization process for monocular thermal odometry. (2) The high-resolution and high-frequency depth data significantly enhances the motion estimation by the known depth. (3) Depth cameras are independent of illumination, making them a suitable choice to aid thermal odometry in poor illumination environments.

It should also be noted that raw thermographic data is represented in a high dynamic range (e.g., 14-bit or 16-bit). Some methods scale thermal images into the 8-bit range to make them compatible with existing vision algorithms. The scaling operation may lead to some problems, such as lower contrast, amplified noise, and photometric inconsistency. Shin et al. [5] proposed an image enhancement method for the thermal domain, with the cost of increasing the noise magnitude. Das et al. [6] introduced an online photometric calibration approach for 8-bit thermal images. However, the calibrated images remain low contrast compared to raw thermographic data.

In this letter, we propose a novel semi-direct thermal-depth odometry method. The proposed method provides a lightweight solution for indoor localization under different illumination (a sample result is shown in Fig. 1). In particular, an NUC handling module is introduced to improve robustness during NUC

Manuscript received 8 November 2022; accepted 19 April 2023. Date of publication 28 April 2023; date of current version 25 May 2023. This letter was recommended for publication by Associate Editor F. Caballero and Editor J. Civera upon evaluation of the reviewers' comments. This work was supported in part by STI 2030-Major Projects under Grant 2021ZD0201403, in part by NSFC under Grant 62088101 Autonomous Intelligent Unmanned Systems, in part by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China under Grant ICT2022B04, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ22F030022. *(Corresponding author: Yu Zhang.)*

Xingxin Chen, Bin He, and Yu Zhang are with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310007, China (e-mail: chenxingxin@zju.edu.cn; binhe@zju.edu.cn; zhangyu80@zju.edu.cn).

Weichen Dai is with the College of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: weichendai@hotmail.com).

Jiajun Jiang is with the Alibaba Group, Hangzhou 310052, China (e-mail: elkulasjiang@zju.edu.cn).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2023.3271510, provided by the authors.

Digital Object Identifier 10.1109/LRA.2023.3271510

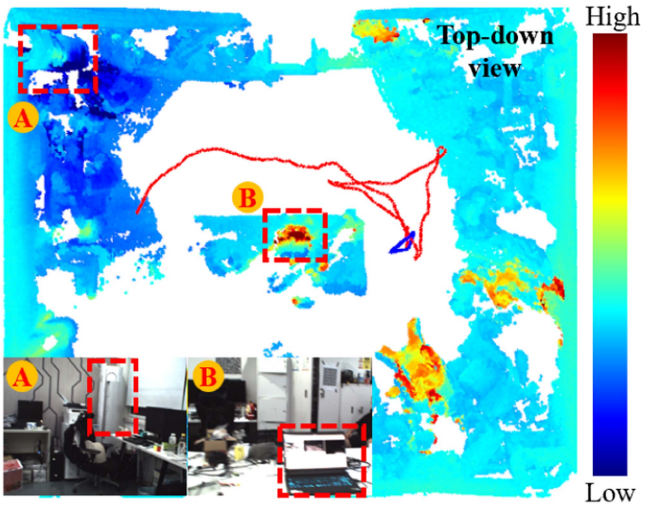


Fig. 1. The proposed thermal-depth odometry estimates the ego-motion (red trajectory) and builds a 3D thermographic map of an office room (top-down view). The point cloud is registered by back-projecting the depth maps using the estimated keyframe poses. No point-cloud-fusion algorithm is applied. The color of the point cloud represents the temperature of the scene. Region A, with an air conditioner, shows a low temperature, and Region B, with a laptop, displays a high temperature.

operation. The relative pose during NUC can be estimated by registering point clouds generated by depth images. We adopt a semi-direct framework in the front-end, fully exploiting raw thermographic data for feature extraction and association. Furthermore, the depth measurements make the tracking more stable and contribute to the tracking recovery after NUC operation. The main contributions of this work are as follows:

- We propose a visual odometry method for a stereo setup of a thermal and a depth camera.
- We propose an NUC handling module that is based on correspondence-based point cloud registration. This module is designed to improve stability in the presence of NUC events. Moreover, we exploit online aligned depth information to enhance tracking robustness, leveraging prior knowledge of scale and 3D structure.
- We conduct experiments on public datasets. The results indicate that the proposed method provides robust pose estimation under different illumination conditions.

II. Related Work

Many VO or Simultaneous Localization and Mapping (SLAM) methods in the visible spectrum have been studied in the literature. Most approaches fall into three categories: feature-based methods [7], direct methods [8], and hybrid (semi-direct) approaches [9]. The semi-direct framework performs direct image alignment to obtain an initial pose before establishing the feature association and refining the pose using reprojection-error-based optimization. Nevertheless, these methods rely on sufficient illumination for maintaining photometric consistency and rich texture. To achieve robust pose estimation under different illumination, thermal infrared cameras have raised high attention in recent odometry studies. The proposed thermal-depth odometry follows the semi-direct pipeline but represents several improvements for thermal imaging systems, including raw thermographic data utilization and an NUC handling module.

Thermal camera-based odometry was first introduced by Vidas and Sridharan [10], in which the GFTT features are tracked using optical flow on re-scaled 8-bit images. Mouats et al. [11] proposed a thermal stereo odometry method combining Fast-Hessian interest points with FREAK descriptors to enhance data association. To improve the performance of pure thermal camera-based solutions, some existing works incorporate thermal cameras with other modalities, such as visual [12], radar [13], or inertial [1]. The inertial sensor is a typical choice since it is invariant with environmental change. Khattak et al. [2] proposed keyframe-based Thermal-Inertial Odometry (TIO) tracking with direct methods. It achieves good performance using raw thermographic data, but the algorithm is sensitive to the convergence of initial depth in the initialization process. With the rapid development of Deep Neural Networks (DNNs), some existing works apply DNNs in thermal-inertial odometry or SLAM. Zhao et al. [14] replaced the conventional feature detection module with a deep-learning-based method to extract high-quality features in thermal images. Jiang et al. [1] designed an optical flow network named ThermalRAFT for feature association, showing robust tracking ability in the thermal domain. Saputra et al. [15] proposed DeepTIO, an end-to-end deep neural network to realize thermal-inertial odometry. Further, they extended DeepTIO to a SLAM system with loop closure [16]. Despite the promising results of deep-learning-based methods, a large amount of training data and the dependence on GPUs still hinder real-time applications.

In past decades, thermal applications with depth sensors have been studied. Chen et al. [17] developed a thermal-LiDAR SLAM approach, which tightly couples edge-based visual odometry and LiDAR odometry. Shin and Kim [18] proposed a thermal SLAM system enhanced by sparse depth information from LiDAR. LiDAR depth information enhances the tracking robustness and scale estimation of odometry systems. The point cloud is initially projected onto the image plane, and a small subset is retained via salient point selection. This approach ensures the association of features and sparse depth, but the quality of features may be compromised due to the projection of points in pixels with low gradients. In contrast, we employ a quadtree strategy [7] to extract the most prominent corners and edge points, resulting in more distinctive features. The work presented in [18] turned off NUC to avoid data interruption. However, the NUC procedure is common among thermal cameras and is required to correct temperature drifts [19]. Instead of turning off NUC, our approach incorporates an NUC handling module to enhance the system's robustness. Depth cameras provide a lightweight, cost-effective solution for depth-enhanced thermal odometry. RGB-D camera-aided thermal SLAM methods mostly focus on 3D thermographic mapping [20], [21], [22]. Vidas et al. [20] proposed a thermal mapping approach, which registers the point clouds generated from the RGB-D camera using Iterative Closest Point (ICP) method [23]. Then the temperature information is projected onto the registered point clouds to build a 3D thermographic model. Zhao et al. [21] fused the thermal and RGB-D information to obtain RGB-DT images and combined the geometry and thermographic errors for pose

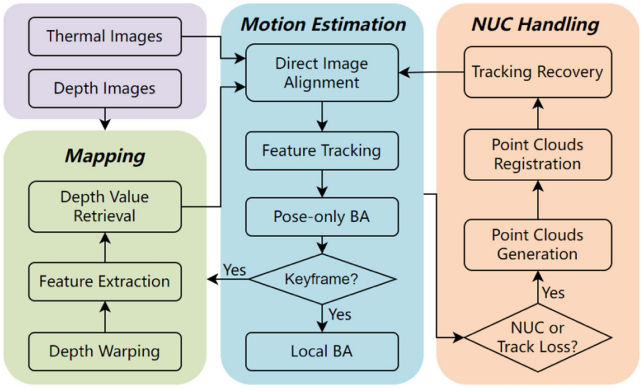


Fig. 2. System overview of the proposed thermal-depth odometry.

estimation. These methods require high computation resources (e.g., GPU) for ICP and dense mapping. The focus of the present work remains on the lightweight localization problem.

For the NUC problem, some methods disabled NUC to avoid image data interruption [17], [18]. Delaune et al. [24] and their relevant work [25] suggested that TIO was not affected by the spatial non-uniformities of thermal images within at least 34 minutes. However, the accumulated temperature drift of thermal images cannot be neglected in long-term navigation. The NUC problem has been a research topic in some previous works. Vidas and Sridharan [10] used SURF descriptors to enhance the feature association before and after an NUC operation. However, the matching performance cannot be guaranteed under significant viewpoint change. Borges and Vidas [26] proposed an NUC manager module to automatically determine when to perform NUC. They considered several factors, such as the current and the predicted rotation. Nevertheless, they neglected camera translation, which may cause significant feature displacement during NUC operation. A feasible solution to overcome the NUC problem is fusing thermal cameras with other sensors [2], [12]. In the present work, we leverage the depth sensor and design an NUC handling module to improve the robustness against NUC operation.

III. Thermal-Depth Odometry

The overview of the proposed method is illustrated in Fig. 2, which consists of three modules: motion estimation, mapping, and NUC handling. The input is 16-bit thermal and depth images. In the motion estimation thread, direct image alignment on raw thermographic data is performed to obtain the initial pose. With the estimated prior pose, the features in the 3D map are tracked in the current frame using KLT tracking. Finally, we apply a reprojection-error-based optimization to refine the pose of the current frame. Once a new keyframe is created, the local BA is launched within the covisibility graph [7], and the keyframe is inserted into the mapping module. Mapping thread proceeds by extracting features from raw thermographic data and assigning depth values from the warped depth images. The NUC handling module is activated upon track loss, designed to overcome the periodic NUC operation.

A. Mapping Module

The mapping module is mainly responsible for feature extraction, depth image warping, and depth value assignment.

1) Feature Extraction: The proposed method utilizes raw thermographic data in feature extraction and association modules. Exploiting raw thermographic data avoids the inconsistent intensity caused by the re-scaling operation. Further, the noise magnitude of 16-bit thermal images is relatively small [2]. These advantages make it more suitable for thermal feature detection and tracking.

We divide the thermal image into cells and extract FAST corners [27]. For the cells where no corner is detected, canny edge points [28] are extracted. To mitigate the effects of noise, we reject the points with gradients lower than a certain threshold. The extracted candidate features and the existing tracked points are distributed using quadtrees [7]. We discard newly detected features in the blocks where tracked points already exist. For the rest of the blocks, the feature with the highest score is selected. The score is generated by Harris response for corners and gradient for edge points.

2) Depth Warping and Assignment: Depth camera readings provide scale information for tracking. To retrieve the depth values of feature points in thermal images, we first warp the depth maps onto the thermal images. We use a linear motion model estimated by the visual odometry to compensate for the time misalignment between the sensors. With the calibrated intrinsic and extrinsic parameters, the depth points from the depth images can be transformed into thermal images by

$$\mathbf{p}_k^t = \pi_t(\mathbf{T}_m \cdot \mathbf{T}_{td} \cdot D(z_k^d) \cdot \pi_d^{-1}(\mathbf{p}_k^d)), \tag{1}$$

where t and d indicate the coordinates of the thermal camera and depth camera, respectively. π_t denotes the projection function of the thermal camera, mapping 3D points in the camera frame to the pixel plane, and π_d^{-1} denotes the inverse projection function of the depth camera, mapping from pixel coordinate to the camera frame. $\mathbf{T}_{td}$ represents the extrinsic transformation matrix from the depth camera to the thermal camera. $\mathbf{T}_m$ is the linear motion transform from depth image timestamp to thermal image timestamp. $D(a) = \text{diag}(a, a, a, 1)$ represents a diagonal matrix. Finally, $\mathbf{p}_k^d$ is the kth pixel coordinate, and z_k^d is its corresponding depth value in the depth camera. $\mathbf{p}_k^t$ is the warped pixel coordinate in the thermal image plane. Its corresponding depth value z_k^t can be obtained by

$$z_k^t = [0010] \cdot \mathbf{T}_m \cdot \mathbf{T}_{td} \cdot D(z_k^d) \cdot \pi_d^{-1}(\mathbf{p}_k^d). \tag{2}$$

Due to the viewpoint difference, the warped depth maps may have a smaller Field of View (FOV) than thermal images. To tackle this problem, we merge the depth maps of history keyframes in the covisibility graph via the relative poses estimated by the visual odometry. The thermal-depth alignment is performed for keyframes only where new features are extracted. To bound the computational complexity, uniform sampling is used when warping the depth points. The warping operation will not significantly increase the computation cost (see Section-IV-E). The warped depth images contain some missing regions, particularly at the boundary of objects. The reasons are the

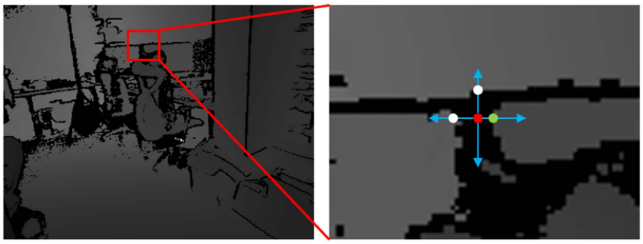


Fig. 3. Illustration of depth value retrieval in miss areas (black areas). Starting from the feature point (red dot), we search the depth values in four directions (blue arrows) in a fixed length and choose the one with the smallest depth value (the green dot associated with the closest computer screen).

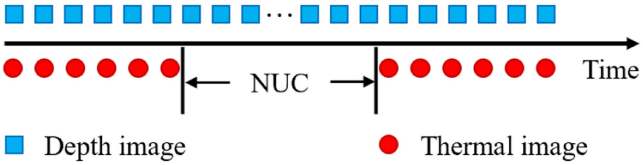


Fig. 4. Illustration of the NUC issue. NUC operation freezes the thermal camera and drops image data.

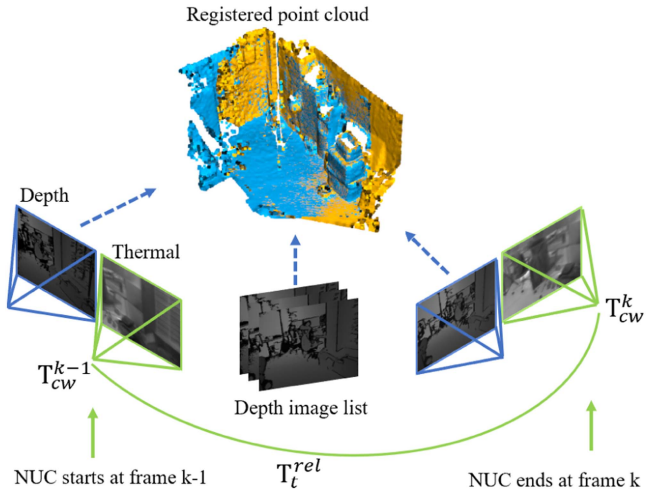


Fig. 5. Illustration of point cloud registration process in NUC handling module. We utilize two or multiple depth images during NUC to generate point clouds and perform point cloud registration.

unique measuring principle of depth cameras and the presence of parallax between the thermal camera and the depth camera. For retrieving depth values in the missing areas, we develop an effective and computation-efficient algorithm, as shown in Fig. 3. Starting from the location of the feature point, our method searches the depth values along the four directions of up, down, left, and right within a certain length. The final depth value is determined as the smallest one of the four directions. The strategy ensures the depth measurements of the missing regions are associated with foreground points.

B. NUC Handling

NUC is one of the major issues in thermal SLAM methods where thermal data interruption causes significant viewpoint change and breaks motion tracking. The NUC issue is illustrated in Fig. 4. In this study, we cope with the NUC issue by leveraging the point cloud registration methods. As shown in Fig. 5, we maintain a list of depth images between two consecutive thermal images. When the camera loses tracking due to NUC occurrence, the NUC handling module is activated. The module first generates point clouds from the depth image list, then derives relative pose by registering the point clouds. For registration methods, classic ICP [23] and its variants [29], [30], [31] rely on a good initial transformation, which requires the point clouds have a significant overlap to use the constant motion model as the initial guess. For that purpose, collecting more point clouds and adopting a scan-to-map registration strategy is a feasible solution. However, the computational time will increase if more depth images are used. Based on the analysis, we choose the correspondence-based method for point cloud registration and select as few depth images as possible. In most cases, only the first and the last point clouds are selected if they sufficiently overlap, and are named source cloud and target cloud, respectively. Specifically, we compute FPFH [32] descriptors for the downsampled point clouds and use bidirectional kNN matching to establish putative correspondences. Given the correspondences, TEASER++ [33] is used to iteratively solve the relative pose between the source and target point clouds. TEASER++ is robust against correspondence outliers and thus is suitable for practical applications. The relative pose $\mathbf{T}_d^{rel}$ calculated by TEASER++ can be transformed into the thermal camera frame by

$$\mathbf{T}_t^{rel} = \mathbf{T}_{td} \cdot \mathbf{T}_d^{rel} \cdot \mathbf{T}_{td}^{-1}, \tag{3}$$

where the subscripts t and d indicate the coordinates of the thermal and depth camera, respectively. $\mathbf{T}_t^{rel}$ is the relative pose in the thermal camera frame. $\mathbf{T}_{td}$ denotes the extrinsic transformation from the depth camera to the thermal camera.

The pose of the kth frame $\mathbf{T}_{cw}^k$ from world frame w to camera frame c can be derived by

$$\mathbf{T}_{cw}^k = \mathbf{T}_t^{rel} \cdot \mathbf{T}_{cw}^{k-1}, \tag{4}$$

where $\mathbf{T}_{cw}^{k-1}$ is the pose of the frame before NUC, as shown in Fig. 5.

After NUC, the current frame is set to keyframe for new feature detection and depth value assignment. With the depth information, the system can initialize the structure quickly, ensuring stable camera tracking. It is noted that the proposed NUC handling module not only helps to overcome the NUC interruption, but also improves robustness in other tracking failure cases such as low-texture.

C. Motion Estimation

Inspired by SVO [9], we adopt a hybrid scheme for motion estimation, which combines the direct approach and the feature-based method. The pose estimation starts with direct image alignment on raw thermographic data. With the relatively reliable depth from the depth camera, the features in the last frame can be projected to the current frame by the initial pose.

Then we minimize the thermographic error as follows:

$$\mathbf{T}_{k,k-1} = \arg\min_{\mathbf{T}_{k,k-1}} \sum_{\mathbf{p}\in\mathbf{P}} \sum_{\mathbf{u}\in\mathbf{\Omega}} \left\| \mathbf{I}_k\left(\mathbf{u}'\right) - \mathbf{I}_{k-1}(\mathbf{u}) \right\|_h , \quad (5)$$

where $\mathbf{T}_{k,k-1}$ represents the relative rigid transform from the last frame $k-1$ to the current frame k. $\mathbf{P}$ is the set of 3D points in the previous frame. $\mathbf{\Omega}$ is the image coordinates in the local patch centered on the feature points. $\mathbf{I}_{k-1}(\mathbf{u})$ denotes the thermographic value of the thermal image at pixel location $\mathbf{u}$. $\|\cdot\|_h$ represents the Huber norm.

Eq. (5) is solved in an iterative Levenberg–Marquardt algorithm. In order to improve the robustness against the fast motion, a multi-scale image pyramid and constant motion model are utilized to process from coarse to fine.

In the second stage, we build frame-to-map feature association using the optical flow method. Through the estimated prior pose, we project all 3D points onto the current frame as the initial position of optical flow tracking. For each 3D point, we choose the keyframe with the closest observation angle with the current frame as the reference keyframe. Based on the initial guess, the 2D feature position is optimized by minimizing the thermographic error of the patch in the current image with respect to the reference patch, which is defined as

$$\mathbf{u}'_i = \arg\min_{\mathbf{u}'_i} \frac{1}{2} \sum_{\mathbf{u}\in\mathbf{\Omega}} w_g \left\| \mathbf{I}_k\left(\mathbf{u}'_i\right) - b_k - \mathbf{A}_i \mathbf{I}_r\left(\mathbf{u}_i\right) + b_r \right\|^2 , \quad (6)$$

where $\mathbf{A}_i$ is an affine warping applied to the ith reference patch. $\mathbf{u}_i$ is the pixel location of map point $\mathbf{p}_i$. b denotes the average thermographic value of the patch, and its subscripts k and r represent the kth frame and the reference keyframe. w_g is the weight function related to thermal gradient [8], which is defined as

$$w_g = \frac{c^2}{c^2 + \|\nabla I\|_2^2}. \quad (7)$$

With the established feature correspondences, we apply a *motion-only* BA to refine the pose of the current frame k, which minimizes the reprojection error as follows:

$$\mathbf{T}_{k,w} = \arg\min_{\mathbf{T}_{k,w}} \frac{1}{2} \sum_i \left\| \mathbf{u}_i - \pi_t\left(\mathbf{T}_{k,w}\mathbf{p}_i\right) \right\|^2 . \quad (8)$$

If the current frame is set to a new keyframe, a local BA is performed within the covisibility graph, jointly optimizing the keyframe poses as well as the observed 3D points.

IV. Experiments

A. Evaluation Setup

1) Dataset: We evaluate the proposed method on public datasets, including the multi-spectral dataset [34] and ViViD++ dataset [35].

The multi-spectral dataset was acquired by a handheld device. We use the sequences recorded in the office room. The sequences were collected under various illumination conditions, including bright, varying, and dim. In addition to the publicly available data, we collected several sequences to thoroughly test the performance under dark illumination and the effectiveness of the NUC handling module. We recorded the data in an office room with the same sensor suit as in [34]. The newly recorded sequences have a prefix *office* in their names.

ViViD++ dataset was collected by a handheld device and a car. For our experiments, we choose 7 handheld sequences to validate the effectiveness of the proposed method. The selected sequences contain different illumination conditions and movement speeds.

2) Evaluation Metrics: We evaluate the accuracy using Root Mean Square (RMS) of Absolute Trajectory Error (ATE) and Relative Pose Error (RPE). Since our method is an odometry system with no loop closure, we turn off the loop closure module of competing SLAM methods to make a fair comparison. We mark the result as fail $(-)$ if the estimated trajectory is too short or divergent.

B. Comparison Against RGB-D Camera-Based Methods

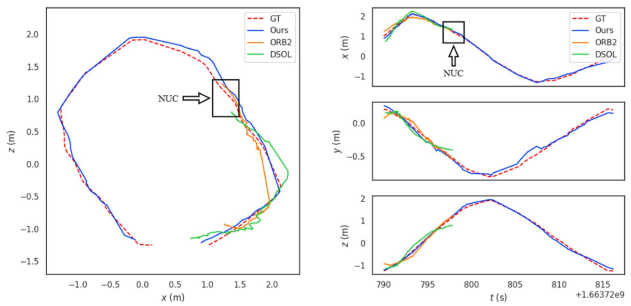
In this subsection, we compare the proposed method against ORB-SLAM2 [7] and DSOL [36], which are leading feature-based and direct RGB-D odometry methods, respectively. All the algorithms are tested on the multi-spectral dataset. The ATE and RPE results are shown in Table I. The upper row of the table indicates two types of input information: RGB and depth images (RGB-D), thermal and depth images (Thermal-D).

As can be seen, the proposed method outperforms the competing algorithms. For RGB-D implementation, ORB-SLAM2 and DSOL fail in varying and dark illumination, while our method can conduct pose estimation in different lighting conditions. The reason is that thermal images are naturally independent of illumination. In bright environments, our method achieves competitive results compared to classic approaches with RGB-D implementation. In some sequences prefixed with *desk*, the proposed method is more accurate, possibly because of the sufficient thermal features in those sequences and the absence of thermally-flat regions. For instance, a static person occupies the majority of the view and provides thermal texture in desk1-xyz-person. In the Thermal-D implementation, ORB-SLAM2 and DSOL become more robust to visible light disturbances. However, the results are still sub-optimal compared to our method. Since DSOL highly relies on photometric consistency between consecutive frames, we re-scale the raw thermal images in a fixed temperature range for DSOL. This re-scaling operation lowers the contrast of thermal images, resulting in a higher trajectory error. Moreover, image interruption during NUC leads to several failures of DSOL. As for ORB-SLAM2, its Thermal-D implementation can provide odometry results in most sequences with the aid of the relocalization module. However, ORB-SLAM2 cannot handle the NUC problem when the camera does not return to its previously visited place (e.g., office-bright-circle and office-dark-circle). The accuracy of ORB-SLAM2 becomes worse in the sequences that contain texture-less or high-noise images. The performance variation of ORB-SLAM2 shows links with the textures and movements. Fig. 6 indicates that the competing methods fail to complete the sequence due to

TABLE I
EVALUATION RESULTS ON MULTI-SPECTRAL DATASET (ATE : METER, T_{RPE} : METER, R_{RPE} : DEGREE)

Sequence	RGB-D						Thermal-D								
	ORB2			DSOL			ORB2			DSOL			OURS		
Bright	ATE	T_{rpe}	R_{rpe}	ATE	T_{rpe}	R_{rpe}	ATE	T_{rpe}	R_{rpe}	ATE	T_{rpe}	R_{rpe}	ATE	T_{rpe}	R_{rpe}
desk1-xyz-person	0.0402	0.0105	0.3668	0.0820	0.0116	0.4217	0.1412	0.0106	0.4404	0.2899	0.0314	0.9023	**0.0154**	**0.0057**	**0.3276**
desk1-halfsphere	0.1061	0.0139	0.8585	0.0654	0.0118	**0.3329**	0.4416	0.1093	5.3256	0.3866	0.0252	1.1847	**0.0582**	**0.0095**	0.4499
desk2-xyz	0.0618	0.0180	0.5873	0.0641	0.0194	0.3991	-	-	-	-	-	-	**0.0375**	**0.0091**	**0.3872**
desk2-halfsphere	0.0759	0.0104	0.6798	**0.0195**	**0.0066**	**0.4011**	0.1525	0.0929	0.9074	-	-	-	0.0288	0.0073	0.5301
office-bright-circle*	**0.0407**	**0.0055**	0.5195	0.0939	0.0127	**0.3913**	-	-	-	-	-	-	0.0838	0.0115	0.7029
office-bright-xyz1*	**0.0139**	**0.0041**	**0.4466**	0.0698	0.0180	0.4784	-	-	-	-	-	-	0.0505	0.0083	0.5107
office-bright-halfsphere1*	**0.0291**	0.0104	**0.2804**	0.0577	0.0149	0.3220	-	-	-	-	-	-	0.0592	**0.0090**	0.2895
office-bright-random1*	**0.0638**	0.0111	0.3391	0.0752	0.0172	0.4085	-	-	-	-	-	-	0.1017	**0.0106**	**0.3336**
Varying															
desk1-xyz-ic	-	-	-	-	-	-	0.2502	0.0991	2.2364	-	-	-	**0.1151**	**0.0111**	**0.4548**
desk1-xyz-ic-person*	-	-	-	-	-	-	0.3490	0.0555	0.5659	-	-	-	**0.0778**	**0.0126**	**0.4053**
desk1-halfsphere-ic*	-	-	-	-	-	-	0.4674	0.0906	1.1653	0.5405	0.1117	1.3380	**0.0514**	**0.0073**	**0.4710**
desk2-xyz-ic*	-	-	-	-	-	-	0.2032	0.1647	2.6294	-	-	-	**0.1136**	**0.0256**	**0.6217**
desk2-halfsphere-ic*	-	-	-	-	-	-	0.1889	0.0297	1.1939	-	-	-	**0.1099**	**0.0165**	**0.5151**
office-vary-xyz1*	-	-	-	-	-	-	-	-	-	-	-	-	**0.0569**	**0.0091**	**0.5325**
office-vary-halfsphere1*	-	-	-	-	-	-	-	-	-	-	-	-	**0.0753**	**0.0117**	**0.4536**
office-vary-random1*	-	-	-	-	-	-	-	-	-	-	-	-	**0.1113**	**0.0083**	**0.2669**
Dark															
office-dark-xyz1	-	-	-	-	-	-	0.3893	0.0431	0.9320	0.4253	0.0650	0.8310	**0.1505**	**0.0127**	**0.3836**
office-dark-xyz2	-	-	-	-	-	-	0.2231	0.0219	0.3255	0.4878	0.0600	0.8243	**0.0806**	**0.0095**	**0.3239**
office-dark-halfsphere1	-	-	-	-	-	-	0.1923	0.0258	0.9604	0.5094	0.0568	0.6561	**0.0972**	**0.0153**	**0.6544**
office-dark-halfsphere2	-	-	-	-	-	-	0.2355	0.0340	1.1233	0.5236	0.1106	0.8291	**0.0685**	**0.0193**	**0.5395**
office-dark-circle*	-	-	-	-	-	-	-	-	-	-	-	-	**0.1025**	**0.0158**	**0.4767**
office-dark-xyz3*	-	-	-	-	-	-	-	-	-	-	-	-	**0.0708**	**0.0149**	**0.3708**
office-dark-halfsphere3*	-	-	-	-	-	-	-	-	-	-	-	-	**0.0816**	**0.0205**	**0.6631**
office-dark-random1*	-	-	-	-	-	-	-	-	-	-	-	-	**0.1426**	**0.0182**	**0.3594**

* Sequence that contains NUC.



(a) (b)

Fig. 6. (a) Estimated trajectory aligned with ground truth in office-dark-circle. (b) Estimated trajectory along each axis. ORB-SLAM2 and DSOL fail to complete the sequence due to NUC (black rectangle).

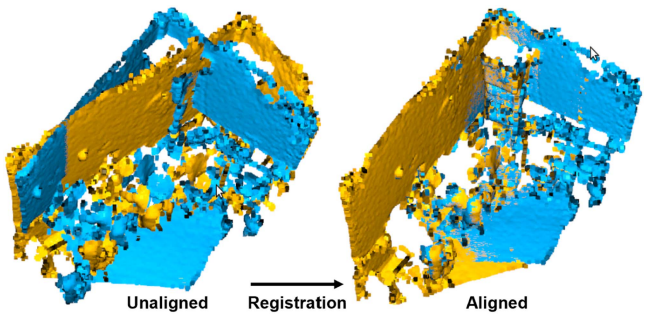


Fig. 7. Point cloud registration results in office-dark-circle (source cloud in orange and target cloud in blue).

the NUC problem. Fig. 7 shows the corresponding point cloud registration result when the NUC occurs.

C. Comparison Against Thermal Odometry Methods

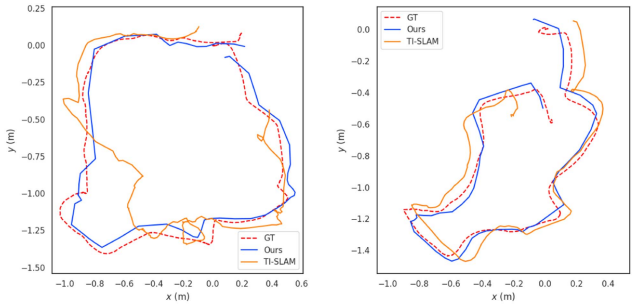
Since there are few open-source thermal-depth odometry methods, we compare the proposed odometry system with state-of-the-art thermal-inertial methods. The competing algorithms include TI-SLAM [1], DeepTIO [15], and xVIO [24]. In particular, DeepTIO is an end-to-end deep thermal-inertial odometry. xVIO is based on Extended Kalman Filter (EKF) and tightly couples inertial data and thermal images. The evaluation is conducted on the ViViD++ dataset. For DeepTIO, the authors fine-tune the pre-trained model on the ViViD++ dataset with the default configuration (e.g., learning rate). The parameters of TI-SLAM and xVIO are tuned to attain the best results from the authors' efforts.

The evaluation results are presented in Table II. The reported values show that the proposed method achieves the best accuracy in most sequences. TI-SLAM provides good results in slow sequences, benefiting from its SVD-based image processing for feature extraction and ThermalRAFT network for feature association. In some sequences, TI-SLAM's rotational RPE is better, which is attributed to the smooth rotation prior provided by IMU. In unstable sequences, TI-SLAM suffers from the initialization problem due to the fast motion. Our method uses depth measurements to initialize the 3D structure, resulting in a

TABLE II
EVALUATION RESULTS ON VIVID++ DATASET (ATE : METER, T_{RPE} : METER, R_{RPE} : DEGREE)

Sequence	xVIO			DeepTIO			TI-SLAM			OURS		
	ATE	T_{rpe}	R_{rpe}	ATE	T_{rpe}	R_{rpe}	ATE	T_{rpe}	R_{rpe}	ATE	T_{rpe}	R_{rpe}
indoor-global-slow*	0.5934	0.0330	0.7217	0.3109	0.0250	1.3010	0.2697	0.0121	**0.3108**	**0.1098**	**0.0083**	0.4274
indoor-local-slow	0.3782	0.0345	0.0591	0.2456	0.0169	0.6235	0.0591	0.0072	**0.3440**	**0.0524**	**0.0061**	0.4745
indoor-varying-slow*	-	-	-	0.4485	0.0255	1.9180	0.1144	0.0132	0.9887	**0.1036**	**0.0115**	**0.9058**
indoor-dark-slow	0.4091	0.0282	0.9630	0.3581	0.0258	1.0105	0.0851	0.0080	0.4685	**0.0834**	**0.0076**	**0.4583**
indoor-global-unstable	-	-	-	0.6784	0.0587	2.4034	-	-	-	**0.3140**	**0.0288**	**1.1213**
indoor-local-unstable	-	-	-	0.3004	0.0179	1.1762	0.4146	0.0227	0.8311	**0.0867**	**0.0089**	**0.6348**
indoor-dark-unstable	-	-	-	0.5676	0.0319	2.1519	-	-	-	**0.1974**	**0.0204**	**1.2986**

* Sequence that contains NUC.



(a) (b)

Fig. 8. Estimated trajectory aligned with ground truth. (a) Trajectory in indoor-global-slow. (b) Trajectory in indoor-varying-slow.

TABLE III
ATE (M) OF DIFFERENT VERSIONS OF OUR METHOD

Sequence	Illumination	Ours Only Depth	Ours Only Thermal	Ours Thermal-Depth
desk1-xyz	bright	0.1875	0.1651	**0.0719**
desk1-xyz-ic-person*	varying	0.0998	-	**0.0778**
desk2-xyz	bright	0.1389	0.1822	**0.0375**
desk2-halfsphere	bright	0.1352	0.0883	**0.0288**
office-bright-xyz2*	bright	0.1327	-	**0.0734**
office-bright-xyz3*	bright	0.1724	-	**0.0656**
office-vary-halfsphere2*	varying	0.1214	-	**0.0699**

* Sequence that contains NUC.

TABLE IV
REAL-TIME PERFORMANCE OF THE MAJOR MODULES (MILLISEC)

Module	Operation	Average Time
NUC Handling	FPFH Descriptors	44.78
	kNN Matching	112.70
	TEASER++ Registration	5.77
	Total Time	163.25
Mapping	Feature Extraction	9.21
	Depth Warping	14.75
	Depth Assignment	0.11
	Total Time	24.07
Motion Estimation	Direct Image Alignment	4.97
	Feature Tracking	2.70
	Pose-only BA	0.63
	Local BA	26.71
	Total Time	35.01

more robust initialization process. xVIO is designed for rescaled 8-bit thermal images, in which the FAST corners and the KLT tracking are affected by the noise and low contrast. For that reason, xVIO has a high failure rate. Moreover, the authors notice that the state estimation of xVIO tends to be divergent during NUC since the visual constraints do not well correct the IMU bias. With the proposed NUC handling module, our method can handle the NUC issue by registering point clouds and hence provide better trajectory accuracy. Fig. 8 depicts the trajectory examples. DeepTIO can produce odometry results where TI-SLAM and xVIO completely fail. It has great potential in the thermally-flat regions. In our experiments, the possible reason limiting the DeepTIO performance is the inadequate training data in ViViD++ handheld sequences. Our method estimates the camera motion using raw thermographic data and depth camera readings, showing better ATE and RPE results. The error slightly increases in the unstable sequences due to higher movement speed, but the proposed method still outperforms the competing approaches.

D. Ablation Study

This subsection is meant to examine the contribution of thermal and depth information. We compare the full version of our method (as *Thermal-Depth*) with two limited versions that use depth images only (as *Only Depth*) or use thermal images only (as *Only Thermal*). We choose several sequences in the multi-spectral dataset for evaluation. For *Only Depth* implementation, we perform the odometry estimation using FPFH descriptors and the TEASER++ method. The *Only Thermal* implementation is a monocular odometry system. The evaluation results in Table III demonstrate that combining thermal and depth information improves system robustness and accuracy. The *Only Thermal* implementation fails to complete the entire sequence if there are NUC operations. The depth images provide prior information for direct image alignment and feature tracking. Thus, the *Thermal-Depth* implementation shows better ATE results than the *Only Thermal* implementation. Moreover, the full version of our method outperforms the *Only Depth* implementation. The main reason is the limited constraints for point cloud registration when utilizing depth cameras, which results in a reduced accuracy of depth-only odometry.

E. Efficiency Analysis

In this subsection, we test the run-time of different components of the proposed system. The evaluation is performed on a laptop with an Intel i7-10870H processor. Table IV shows the

average computation time in the indoor-varying-slow sequence in ViViD++ dataset. The time cost of the motion estimation thread is 35.01 ms (28.56 FPS) on average. It is noted that the local BA is performed only after the keyframe creation. With the uniform sampling and the efficient depth retrieving strategy, the mapping thread only requires 24.07 ms on average. The NUC handling module consumes 163.25 ms each time with the downsampled point clouds (voxel size is set to 0.1 m). Since it is only activated when the NUC occurs, the overall time consumption will not significantly increase. Based on the analysis above, the proposed odometry system can permit real-time applications with a standard CPU.

V. Conclusion

In this work, we present a semi-direct thermal-depth odometry method enhanced by depth measurements from a depth camera. The proposed NUC handling module overcomes the periodic interruption of thermal cameras. The utilization of the raw thermographic and depth information improves the motion estimation accuracy in the thermal domain. The experiment results demonstrate that our method outperforms the thermal image-based and RGB-D camera-based alternatives. Furthermore, the proposed approach enables lightweight autonomous systems to navigate under different illumination. Nevertheless, the proposed algorithm may fall short in environments with low thermal contrast, such as planar walls, due to the absence of thermal features. In future work, we intend to explore potential solutions in such circumstances. Additionally, the tasks of incorporating a loop closure module and designing a thermal image processing algorithm are topics for future research.

References

[1] J. Jiang, X. Chen, W. Dai, Z. Gao, and Y. Zhang, "Thermal-inertial SLAM for the environments with challenging illumination," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 8767–8774, Oct. 2022.

[2] S. Khattak, C. Papachristos, and K. Alexis, "Keyframe-based thermal-inertial odometry," *J. Field Robot.*, vol. 37, no. 4, pp. 552–579, 2020.

[3] W. Liu et al., "TLIO: Tight learned inertial odometry," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 5653–5660, Oct. 2020.

[4] J. Steinbrener, C. Brommer, T. Jantos, A. Fornasier, and S. Weiss, "Improved state propagation through AI-based pre-processing and down-sampling of high-speed inertial data," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 6084–6090.

[5] U. Shin, K. Lee, B.-U. Lee, and I. S. Kweon, "Maximizing self-supervision from thermal image for effective self-supervised learning of depth and ego-motion," *IEEE Robot. Automat. Lett.*, vol. 7, no. 3, pp. 7771–7778, Jul. 2022.

[6] M. P. Das, L. Matthies, and S. Daftry, "Online photometric calibration of automatic gain thermal infrared cameras," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 2453–2460, Apr. 2021.

[7] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[8] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, Mar. 2018.

[9] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect visual odometry for monocular and multicamera systems," *IEEE Trans. Robot.*, vol. 33, no. 2, pp. 249–265, Apr. 2017.

[10] S. Vidas and S. Sridharan, "Hand-held monocular SLAM in thermal-infrared," in *Proc. IEEE 12th Int. Conf. Control Automat. Robot. Vis.*, 2012, pp. 859–864.

[11] T. Mouats, N. Aouf, L. Chermak, and M. A. Richardson, "Thermal stereo odometry for UAVs," *IEEE Sensors J.*, vol. 15, no. 11, pp. 6335–6347, Nov. 2015.

[12] W. Dai, Y. Zhang, D. Sun, N. Hovakimyan, and P. Li, "Multi-spectral visual odometry without explicit stereo matching," in *Proc. IEEE Int. Conf. 3D Vis.*, 2019, pp. 443–452.

[13] C. Doer and G. F. Trommer, "Radar visual inertial odometry and radar thermal inertial odometry: Robust navigation even in challenging visual conditions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 331–338.

[14] S. Zhao, P. Wang, H. Zhang, Z. Fang, and S. Scherer, "TP-TIO: A robust thermal-inertial odometry with deep thermalpoint," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 4505–4512.

[15] M. R. U. Saputra et al., "DeepTIO: A deep thermal-inertial odometry with visual hallucination," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1672–1679, Apr. 2020.

[16] 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 Trans. Robot.*, vol. 38, no. 3, pp. 1875–1893, Jun. 2022.

[17] W. Chen, Y. Wang, H. Chen, and Y. Liu, "EIL-SLAM: Depth-enhanced edge-based infrared-LiDAR SLAM," *J. Field Robot.*, vol. 39, no. 2, pp. 117–130, 2022.

[18] Y.-S. Shin and A. Kim, "Sparse depth enhanced direct thermal-infrared SLAM beyond the visible spectrum," *IEEE Robot. Automat. Lett.*, vol. 4, no. 3, pp. 2918–2925, Jul. 2019.

[19] M. Vollmer and K.-P. Möllmann, *Infrared Thermal Imaging: Fundamentals, Research and Applications*. Hoboken, NJ, USA: Wiley, 2017.

[20] S. Vidas, P. Moghadam, and M. Bosse, "3D thermal mapping of building interiors using an RGB-D and thermal camera," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2013, pp. 2311–2318.

[21] S. Zhao, Z. Fang, and S. Wen, "A real-time handheld 3D temperature field reconstruction system," in *Proc. IEEE 7th Annu. Int. Conf. CYBER Technol. Automat., Control, Intell. Syst.*, 2017, pp. 289–294.

[22] Y. Cao et al., "Depth and thermal sensor fusion to enhance 3D thermographic reconstruction," *Opt. Exp.*, vol. 26, no. 7, pp. 8179–8193, 2018.

[23] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," *Proc. SPIE*, vol. 1611, pp. 586–606, 1992.

[24] J. Delaune, R. Hewitt, L. Lytle, C. Sorice, R. Thakker, and L. Matthies, "Thermal-inertial odometry for autonomous flight throughout the night," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 1122–1128.

[25] V. Polizzi, R. Hewitt, J. Hidalgo-Carrió, J. Delaune, and D. Scaramuzza, "Data-efficient collaborative decentralized thermal-inertial odometry," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 10681–10688, Oct. 2022.

[26] P. V. K. Borges and S. Vidas, "Practical infrared visual odometry," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2205–2213, Aug. 2016.

[27] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, Jan. 2010.

[28] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[29] J. Serafin and G. Grisetti, "NICP: Dense normal based point cloud registration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 742–749.

[30] A. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Proc. Robot.: Sci. Syst.*, Seattle, WA, 2009, vol. 2, Art. no. 435.

[31] J. Wang, M. Xu, F. Foroughi, D. Dai, and Z. Chen, "FasterGICP: Acceptance-rejection sampling based 3D LiDAR odometry," *IEEE Robot. Automat. Lett.*, vol. 7, no. 1, pp. 255–262, Jan. 2022.

[32] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3212–3217.

[33] H. Yang, J. Shi, and L. Carlone, "TEASER: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, Apr. 2021.

[34] W. Dai, Y. Zhang, S. Chen, D. Sun, and D. Kong, "A multi-spectral dataset for evaluating motion estimation systems," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 5560–5566.

[35] A. Lee, Y. Cho, Y.-S. Shin, A. Kim, and H. Myung, "ViViD++ : Vision for visibility dataset," *IEEE Robot. Automat. Lett.*, vol. 7, no. 3, pp. 6282–6289, Jul. 2022.

[36] C. Qu, S. S. Shivakumar, I. D. Miller, and C. J. Taylor, "DSOL: A fast direct sparse odometry scheme," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 10587–10594.