*remote sensing*

MDPI

*Article*

# A Base-Map-Guided Global Localization Solution for Heterogeneous Robots Using a Co-View Context Descriptor

**Xuzhe Duan [1,2], Meng Wu [1], Chao Xiong [1], Qingwu Hu [2] and Pengcheng Zhao [2,*]**

[1] State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China; duanxz@whu.edu.cn (X.D.)
[2] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430070, China; huqw@whu.edu.cn
* Correspondence: pengcheng.zhao@whu.edu.cn; Tel.: +86-159-7200-3670

**Abstract:** With the continuous advancement of autonomous driving technology, an increasing number of high-definition (HD) maps have been generated and stored in geospatial databases. These HD maps can provide strong localization support for mobile robots equipped with light detection and ranging (LiDAR) sensors. However, the global localization of heterogeneous robots under complex environments remains challenging. Most of the existing point cloud global localization methods perform poorly due to the different perspective views of heterogeneous robots. Leveraging existing HD maps, this paper proposes a base-map-guided heterogeneous robots localization solution. A novel co-view context descriptor with rotational invariance is developed to represent the characteristics of heterogeneous point clouds in a unified manner. The pre-set base map is divided into virtual scans, each of which generates a candidate co-view context descriptor. These descriptors are assigned to robots before operations. By matching the query co-view context descriptors of a working robot with the assigned candidate descriptors, the coarse localization is achieved. Finally, the refined localization is done through point cloud registration. The proposed solution can be applied to both single-robot and multi-robot global localization scenarios, especially when communication is impaired. The heterogeneous datasets used for the experiments cover both indoor and outdoor scenarios, utilizing various scanning modes. The average rotation and translation errors are within 1° and 0.30 m, indicating the proposed solution can provide reliable localization support despite communication failures, even across heterogeneous robots.

**Keywords:** global localization; heterogeneous data registration; LiDAR point cloud; multi robots



**Citation:** Duan, X.; Wu, M.; Xiong, C.; Hu, Q.; Zhao, P. A Base-Map-Guided Global Localization Solution for Heterogeneous Robots Using a Co-View Context Descriptor. *Remote Sens.* **2024**, *16*, 4027. https://doi.org/10.3390/rs16214027

Academic Editors: Wanshou Jiang, San Jiang, Duojie Weng and Jianchen Liu

Received: 19 September 2024
Revised: 28 October 2024
Accepted: 28 October 2024
Published: 30 October 2024



## 1. Introduction

The knowledge of one's own location is crucial for all mobile robot applications. For this reason, localization is a core function of mobile robots. Mobile robots can determine their location in outdoor environments with good global navigation satellite system (GNSS) signals. However, the quality of data heavily influences the performance of GNSS, making it unsuitable for use in areas where satellite signals are obstructed, such as indoors, dense urban environments, or forests. In GNSS-denied environments, mobile robots often rely on external sensors such as cameras and LiDAR to estimate their position and orientation. This approach is known as simultaneous localization and mapping (SLAM). LiDAR sensors provide distance measurements by emitting and receiving light. These measurements are more resilient to illumination and appearance changes than visual images from a camera, making LiDAR very popular in the mobile robotics community. Typical applications for LiDAR localization include laser odometry calculation, loop closure detection, and collaborative SLAM, among others.

The laser odometry aims to calculate the relative transformation between LiDAR scans. LOAM [1] is a typical laser odometry method that matches point features to edge/planar to find correspondences between scans. The features are extracted by calculating the

roughness of the point in its local region. This approach has a profound impact and is still adopted by today's state-of-the-art LiDAR SLAM methods [2,3]. However, odometry-based localization is an incremental process, which means that errors in scan-matching will inevitably accumulate in the overall localization results. The loop closure detection is another important application of global LiDAR localization. It determines if the robot has returned to a previously visited location. This module computes the similarity between current and previous data through the abstracted description of laser scans [4–7]. If a loop closure is detected, the relative transformation between the current and the previous locations is computed and participates in the graph optimization to eliminate cumulative errors [8,9].

With the increasing maturity of the single-robot LiDAR localization, more researchers have begun to explore the possibility of localizing multi-robot systems [10]. Multiple partially overlapping maps can be generated from multi-robot systems using incremental SLAM or other mapping techniques. These maps are utilized for mutual localization between robots, also known as multi-robot localization or cross-robot localization. In this case, robots build local maps that are typically defined with respect to the local coordinate system of each single robot. Recovering the global position of each local map is crucial in multi-robot localization tasks.

Based on the operating modes, existing multi-robot localization methods fall into two categories: online processing and offline processing. Online processing, also known as collaborative SLAM, simultaneously detects whether robots meet each other at a certain point known as a rendezvous [11]. The scans captured at the rendezvous region serve as the inter-robot loop closure to calculate the relative transformations between robots. These constraints are then added into a graph optimization framework to optimize the global location of multi-robot systems. This approach is appropriate for applications with strong real-time requirements, such as military autonomous collaboration, robot navigation, and search and rescue (SAR) operations [12–17]. However, ensuring communications between multi-robot systems, which online processing depends on, can be challenging in practical situations. Communication problems, such as delays, interruptions, hidden signals, and out-of-sequence packets, significantly impact the performance of real-time collaborative SLAM algorithms [18]. Offline processing applies the LiDAR place-recognition method to realize the global localization of multi-robot systems. Place recognition is enabled by point cloud registration. Taking one point cloud as a reference, the registration algorithm first estimates the coarse location of the other point cloud based on the correspondence. It then performs the fine alignment using the iterative closest point (ICP) algorithm [19] and its variants [20–22] to refine the localization result. This approach can provide more accurate localization results without relying on communication conditions. It is suitable for applications with lower real-time demands, such as surveys and 3D scene reconstruction [23,24].

With increasing research efforts focused on HD map development in autonomous driving, a growing amount of point cloud data is being acquired and stored in geographic databases [25]. The fusion of these heterogeneous point clouds (e.g., the aerial-based and the ground-based laser scanning point cloud, or the point clouds acquired from different scanning modes on the ground) can effectively compensate for deficiencies in different scanning modes, expand the coverage of the point cloud, and enhance the representation of the scene. However, global localization, which is the basis of heterogeneous data fusion, is challenging due to the diverse characteristics of heterogeneous point clouds, such as scanning perspectives, ranges, and resolutions.

This paper proposes a heterogeneous robots localization solution based on a pre-set base map. This idea of localization by using a pre-set map originates from [26], a study that leverages a 3D map for image-based localization. In our work, mobile robots provide local scans, and the pre-set base map is used as the reference scan. A descriptor with rotational invariance is developed to explore co-visible regions between local scans and the reference scan. The result of descriptor matching is adopted for coarse localization, followed by refinement using the trimmed ICP [20] algorithm. The proposed solution can

be used for the global localization of both single mobile robots and multiple mobile robots in environments with poor communication.

The principal innovations and contributions of this research are summarized below:

- A base-map-guided LiDAR localization solution for heterogeneous robots is proposed. In this case, a pre-set base map is utilized to localize local scans captured by mobile robots. This solution can be applied to both online and offline global LiDAR localization. It is particularly practical in environments where a rendezvous is difficult to form or where communication is not available.
- A novel co-view context descriptor that can detect co-visible regions of heterogeneous point clouds is developed. This discriminative descriptor takes into account the height and density information of the points and extracts co-visible regions for both horizontal and vertical cases. Descriptor matching enables the estimation of a coarse transformation between the local scan and the pre-set base map, facilitating the global localization of mobile robots.
- Three heterogeneous experimental datasets are elaborated to validate the effectiveness of the proposed solution. The base maps are captured from aerial and ground-based platforms, whereas the local scans are derived from LiDAR SLAM and single-station laser scanning. The results indicate that the proposed solution can be implemented either online or offline to localize both homogeneous and heterogeneous point clouds in various environments.

The rest of the paper is organized as follows: Section 2 describes the related works in the field; Section 3 outlines the specifics of global localization utilizing the proposed co-view context descriptor; Section 4 presents the comprehensive experiments and their corresponding results; Section 5 summarizes and concludes this work.

## 2. Related Works

### *2.1. LiDAR Place Recognition*

The basic idea of place recognition is to retrieve the locations with the highest probability based on the global similarity between the LiDAR point cloud and the given map. It has gone through a process from the handcrafted descriptors to the end-to-end deep learning.

The handcrafted descriptors should exhibit discriminative characteristics for different places but maintain similarity for places close to each other. LiDAR Iris [4] encodes the height information of a 3D point cloud into a binary LiDAR-Iris image and transforms it into the Fourier domain to achieve rotational invariance. This approach makes a highly compressed representation of the 3D point cloud, which is specifically effective for nearest neighbor search in place retrieval but cannot be used in geometric pose estimation. Contour context [5] extracts bird's-eye-view (BEV) contours and leverages them to encode local information for place recognition and pose estimation. Scan context [6] projects a 3D point cloud onto a 2D plane and divides the points into azimuthal and radial bins. A compact global feature with rotational invariance is then constructed, facilitating both place recognition and yaw estimation. The stable triangle descriptor [7] maintains a hash table as the global descriptor and accomplishes place recognition through voting on the triangles stored in the table. The pose with six degrees of freedom (DoF) is then estimated using the singular value decomposition (SVD) method. Based on the descriptor construction methods, Wu et al. [27] proposed a hierarchical framework named HL-MRF to hierarchically localize a large number of multi-view scans efficiently and robustly.

The above methods accomplish localization through two phases: place retrieval and pose estimation. With the maturity of deep learning, some researchers propose to directly regress the global pose of a robot in an end-to-end manner. Wang et al. [28] introduce a learning-based approach called PointLoc for LiDAR global pose estimation. The backbone consists of an attention-aided PointNet-style architecture [29]. Luo et al. [30] propose a rotation-invariant network called BEVPlace. The LiDAR point cloud is represented by a BEV image, from which rotation-equivariant local features are extracted using group convolution. The location of the query point cloud is estimated based on the distance

between BEV features. Wang et al. [31] use a neural network to estimate the overlap between scan pairs and construct a sparse but reliable pose graph to localize multi-view point clouds. These end-to-end approaches are entirely data-driven and eliminate the need for traditional pose estimation processing. However, their interpretability and generalization ability still require improvement.

Some researchers focus on the place recognition of heterogeneous point clouds, especially for the airborne laser scanning (ALS) point cloud and the terrestrial laser scanning (TLS) point cloud. Yang et al. [32] extract building outline features from both the ALS and TLS point clouds and use these feature correspondences to estimate the transformation between two point clouds. Avidar et al. [33] propose a local-to-global registration method based on a viewpoint descriptor dictionary. It utilizes the phase correlation of the range images to rapidly find plausible transformations from the local TLS point cloud to the global ALS point cloud. Liang et al. [34] propose a skyline context descriptor to localize TLS point clouds in ALS point clouds. This approach addresses the challenge posed by different perspectives and resolutions of heterogeneous point clouds. Xu et al. [35] address the viewpoint transformation problem by constructing virtual scans and introduce a polar grid descriptor that incorporates point cloud height information. Based on the similarity of descriptors, the place recognition from SLAM point clouds to the TLS map is conducted.

### *2.2. Collaborative LiDAR SLAM*

In collaborative LiDAR SLAM, the robots participating in the mission contribute to one global map to jointly reconstruct the environment. In general, collaborative LiDAR SLAM algorithms are built on top of single-robot SLAM algorithms. The system architectures for multiple robots can be categorized into centralized and distributed.

Centralized systems leave the complex computing to a central server. As the current state-of-the-art collaborative LiDAR SLAM algorithms, LAMP [36] and LAMP 2.0 [37] entail the direct transfer of local laser maps generated by each robot to a central server for map fusion. This process demands significant computational resources on the server and relies heavily on communication performance. CoLRIO [38] is a LiDAR-inertial-based centralized system. It effectively allocates computationally intensive tasks to the central server, thus alleviating the burden for individual robots. COHORT [39] groups heterogeneous robots, including legged and aerial platforms, for autonomously exploring the subterranean environment. Each robot of the team shares submaps to a centralized location for global mapping and mission scheduling. He et al. [40] tightly couple the image and range measurements on each robot and leverage neural networks to extract descriptors for ground and aerial collaborative mapping. Centralized systems can provide significant benefits in terms of accuracy and data management convenience. However, they often require stable server connections and are susceptible to the single point of failure.

Distributed systems attempt to alleviate connection limitations by eliminating the dependence on the centralized server. DOOR-SLAM [41] is a peer-to-peer-based distributed system. It uses NetVLAD descriptors [42] for place recognition and the distributed Gauss–Seidel algorithm [43] for distributed pose graph optimization. Additionally, the pairwise consistency maximization technique (PCM) [44] is performed to identify and reject outlier inter-robot loop closures. Following the architecture of DOOR-SLAM, DiSCo-SLAM [45] expands the scan context [6] to detect inter-loop closures and introduces a two-stage global–local graph optimization framework. DCL-SLAM [46] proposes a distributed loop closure framework that operates without the need for exchanging all raw or feature points, enhancing adaptability to scenarios with limited bandwidth and communication range. Xu et al. [47] develop a system that comprises multiple unmanned ground vehicles (UGV), each of which carries multimodal LiDAR sensors. The UGVs exchange submaps and relative spatial exploration status through a mesh network. Distributed systems can effectively enhance the availability, yet they still necessitate robots to rendezvous or communicate with each other at some juncture to exchange data.

To summarize, the current LiDAR-based multi-robot system localization methods still face challenges. First, in both online collaborative LiDAR SLAM and offline LiDAR place recognition, the failure of a single robot's localization can cause significant drift for all associated robots. Moreover, most widely-used LiDAR localization solutions are designed for homogeneous robot systems, with limited research on heterogeneous robot systems. Lastly, whether the system is centralized or distributed, there are inherent requirements for communication conditions.

## 3. Materials and Methods

### *3.1. System Overview*

In this section, we present a detailed explanation of the base-map-guided LiDAR localization solution and the co-view context descriptor. The workflow is illustrated in Figure 1. The pre-set base map, serving as a reference scan for mobile robots, can be obtained using ALS or other mapping techniques, whereas the local point clouds of each single robot can be captured using the LiDAR SLAM or single-station laser scanning methods. To begin with, we propose the concept of a virtual scan, including the virtual reference scan (VRS) and the virtual local scan (VLS). They are constructed for both the pre-set base map and the local point clouds and are considered as the processing unit for global localization. The height and density information of the virtual scan points is then leveraged to detect horizontal and vertical co-visible regions and combined to generate the co-view context descriptor. The VRS generates the candidate descriptor, whereas the VLS generates the query descriptor. Subsequently, the similarity between candidate and query descriptors is calculated. This process aids in determining the best matching pair of descriptors, thereby facilitating the estimation of the coarse transformation. Finally, the transformation is refined using the trimmed ICP algorithm to achieve precise global localization for heterogeneous point clouds. Each robot is restored to a coordinate frame that is unified with the base map.

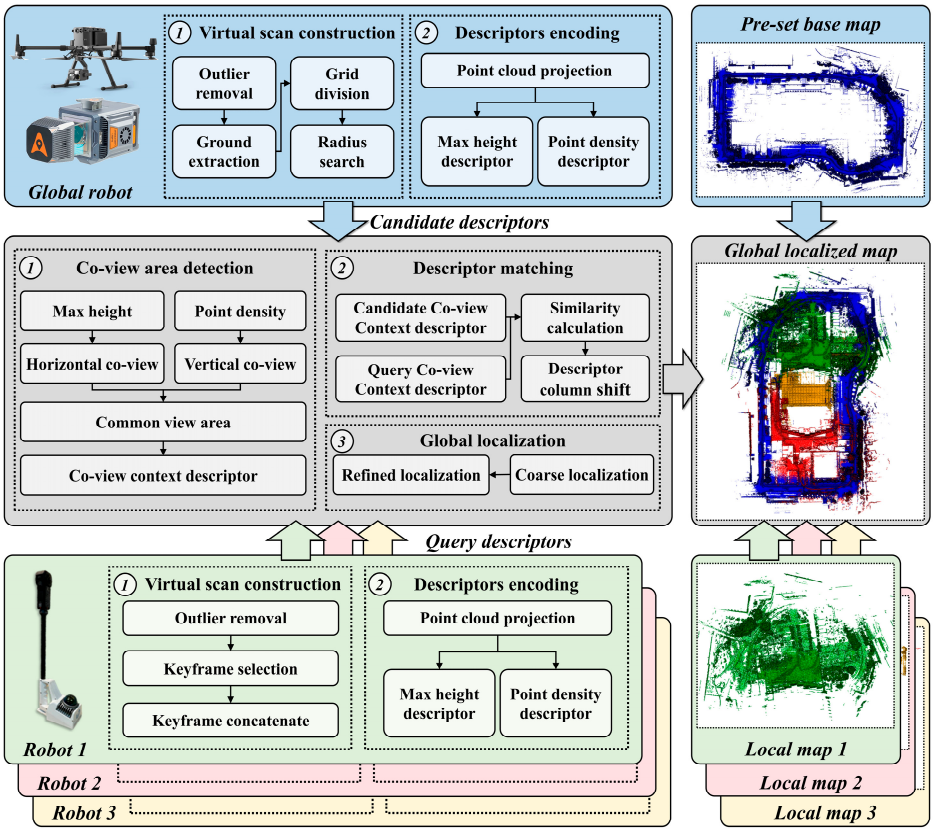


**Figure 1.** The workflow of the base-map-guided global LiDAR localization solution.

### *3.2. Virtual Scan Construction*

Before constructing virtual scans, it is necessary to define a global coordinate frame G, which will serve as the reference for the subsequent point cloud localization. In this paper, we directly set the coordinate system of the base map as the global reference G.

Outlier removal is a preprocessing step aimed at eliminating noise measurements from the original point cloud. In our solution, a simple yet effective statistical outlier removal (SOR) algorithm [48] is employed for noise filtering. A point cloud voxel downsample step is followed to eliminate the effects of varying densities between scans. Virtual scans are subsequentially constructed for both the reference and local scans utilizing two distinct strategies. Each virtual scan serves as an individual processing unit for descriptor generation and matching. Dividing the complete point cloud in this manner addresses challenges associated with matching low-overlap point clouds. The specific implementation is described in detail below.

#### 3.2.1. Virtual Reference Scan Construction

The pre-set base map can be captured through various techniques, including ALS, mobile laser scanning (MLS), and more. Without loss of generality, this paper proposes a VRS construction approach that is applicable to various acquisition techniques.

The movement of robots primarily occurs on the ground. Based on this assumption, our approach concentrates the virtual reference stations on the ground to narrow the search space. Therefore, the first step is to extract ground points from the reference scan using the cloth simulation filtering (CSF) algorithm [49], as shown in Figure 2. The CSF algorithm simulates a rigid cloth falling from above onto the surface of an inverted point cloud. The location of the cloth nodes can be determined by analyzing the interactions between the nodes and the corresponding points, thereby generating an approximation of the ground surface. Based on the approximate surface, ground points can be extracted from the original point cloud.

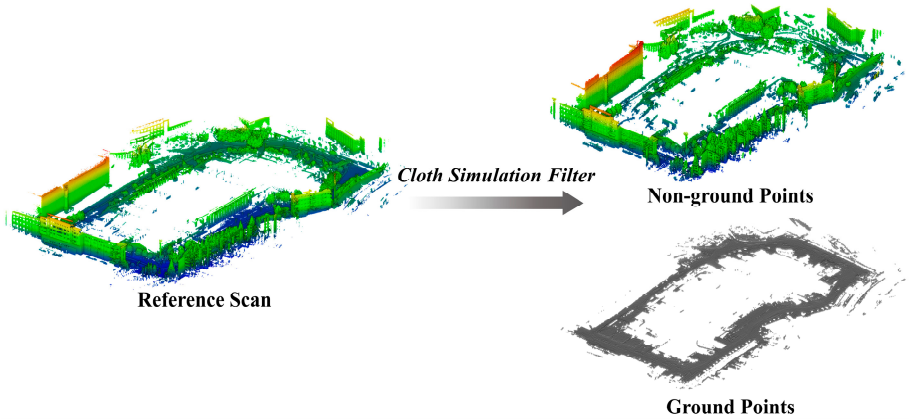


**Figure 2.** Extracting ground points from the reference scan.

Ground points are then divided into grids with a side length of $L$. Each grid's center serves as a virtual reference station, representing the potential locations of the local point clouds. The neighboring points of radius $r_{\text{ref}}$ around each virtual station are searched to form a VRS, as shown in Figure 3. The points of each VRS are translated to the origin of global reference G by $\boldsymbol{t}^{\mathrm{R}} = \left[x^{\mathrm{R}}, y^{\mathrm{R}}, z^{\mathrm{R}}\right]^{\mathbf{T}}$. This represents the translation of the grid's center $\boldsymbol{p}_{\mathrm{C}}^{\mathrm{R}} = \left[x^{\mathrm{R}}, y^{\mathrm{R}}, z^{\mathrm{R}}\right]^{\mathbf{T}}$ to the global coordinate system's origin $\boldsymbol{p}_{\mathrm{O}}^{\mathrm{G}} = [0,\ 0,\ 0]^{\mathbf{T}}$.

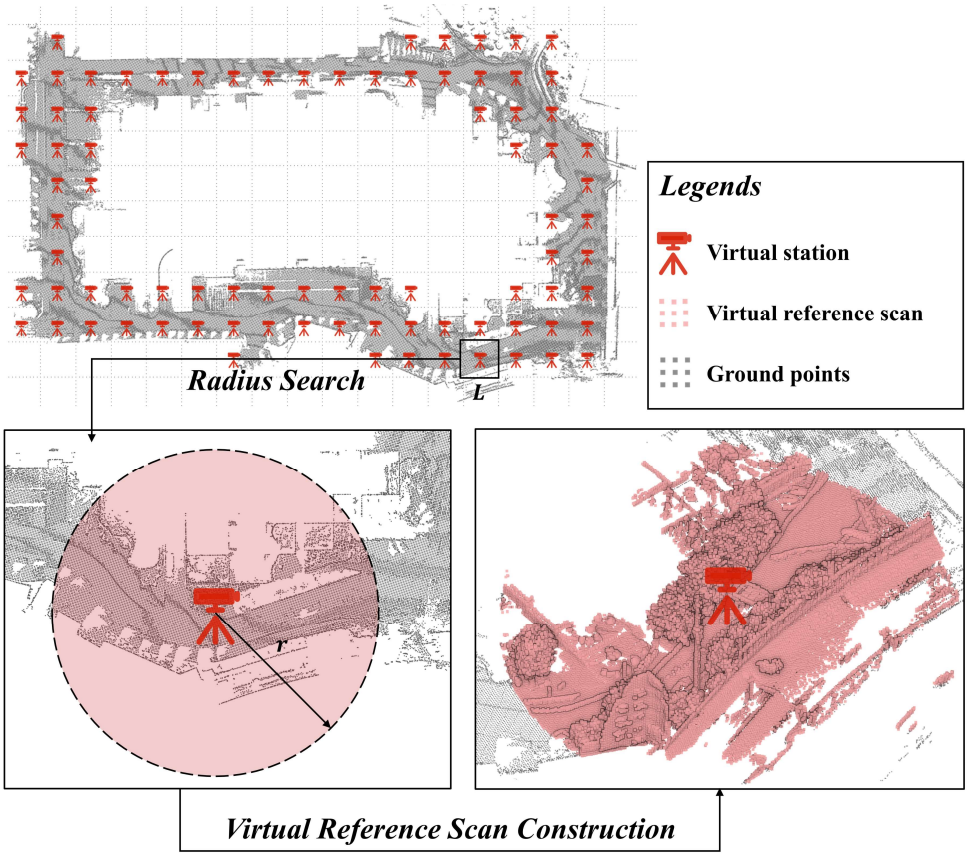


**Figure 3.** The construction of the virtual reference scan.

#### 3.2.2. Virtual Local Scan Construction

Although it is feasible to determine the VLS location from ground points using the approach described above, we find a more straightforward way for mobile robots. The trajectories computed by LiDAR SLAM or other techniques can significantly reduce the search space. It is worth discussing the granularity at which a single robot participates in processing. Matching VRSs with each individual local scan frame is computationally intractable and unnecessary, while matching VRSs with the entire local scan map may result in a low success rate due to limited overlapping regions. To mitigate this issue, we adopt the concept of the keyframe technique, which is widely employed in the visual SLAM field [50,51].

For keyframe selection, we use a simple heuristic: a LiDAR frame is selected as a keyframe when the change in robot position exceeds a user-defined threshold in comparison to the previous position. The scans between two keyframes are accumulated to construct a VLS, which serves as the processing unit for matching with VRSs. The keyframe selection and virtual scan construction process of local scans are shown in Figure 4. The points of each VLS are translated to the origin of global reference G by $\boldsymbol{t}^{\mathrm{L}} = \left[x^{\mathrm{L}}, y^{\mathrm{L}}, z^{\mathrm{L}}\right]^{\mathbf{T}}$. This represents the translation of the keyframe's center $\boldsymbol{p}_{\mathrm{C}}^{\mathrm{L}} = \left[x^{\mathrm{L}}, y^{\mathrm{L}}, z^{\mathrm{L}}\right]^{\mathbf{T}}$ to the global coordinate system's origin $\boldsymbol{p}_{\mathrm{O}}^{\mathrm{G}} = [0,\ 0,\ 0]^{\mathbf{T}}$. Unlike the SLAM keyframe selection strategy, which considers both positional and rotational changes, our approach only relies on positional changes for judgment. This is due to the rotational invariance of the descriptor introduced afterward. Generally, the distance threshold $\sigma_{\mathrm{d}}$ is set to be equal to the grid size $L$ to ensure the same sampling interval.

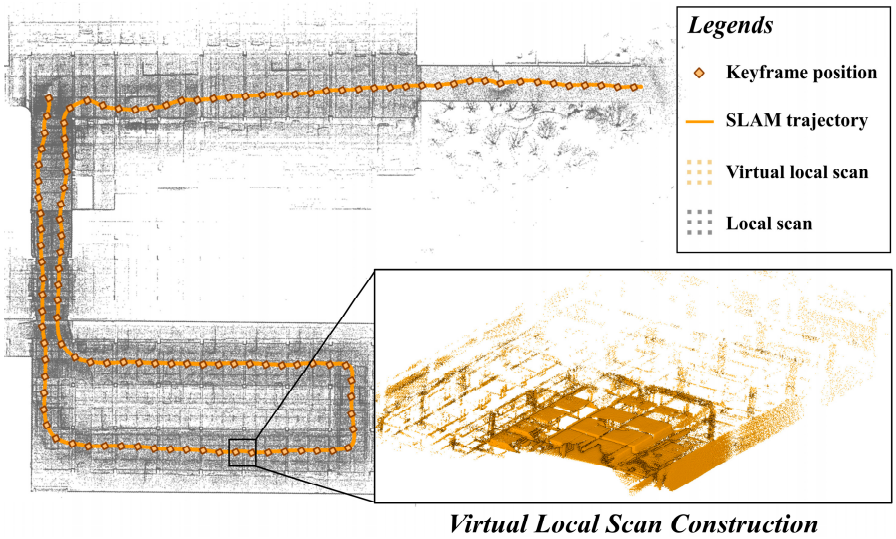


**Figure 4.** The construction of the virtual local scan.

### *3.3. Co-View Context Descriptor*

Inspired by the rotational-invariant scan context descriptor [6] and skyline context descriptor [34], this paper develops the co-view context descriptor to represent the characteristics of heterogeneous laser scans in a unified manner. Following the matching results of descriptors, the global localization of robots is accomplished through a two-phase strategy. The generation and matching processes of the co-view context descriptor are depicted in Figure 5.

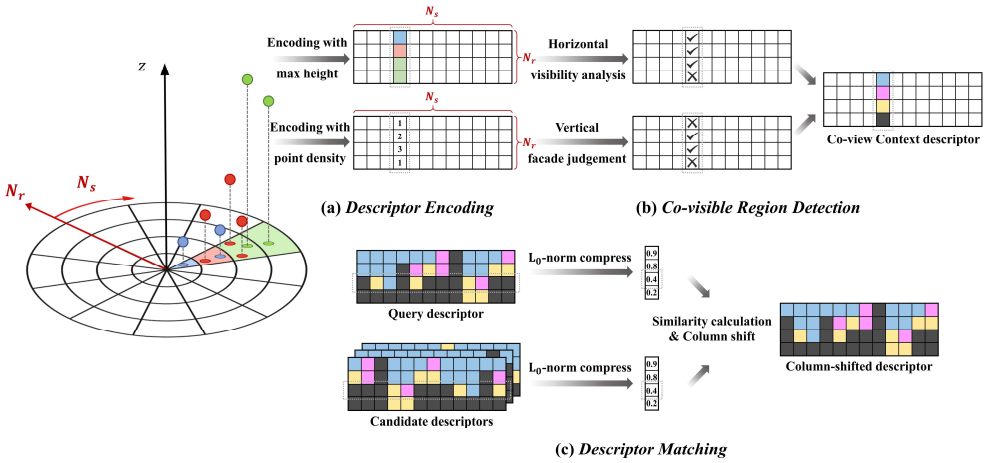


**Figure 5.** The schematic diagram of co-view context descriptor.

#### 3.3.1. Descriptor Encoding

Figure 5a illustrates the encoding process of the descriptors. The virtual scan is projected onto a 2D plane, where points within a range of $r_{\max}$ are mapped into a polar grid with $N_\mathrm{r}$ rings and $N_\mathrm{s}$ sectors. Each bin of the grids employs a value to encode the feature of the points contained within it, such as maximum height, maximum intensity, point density, etc. The polar grid is then transformed into an $N_\mathrm{r} \times N_\mathrm{s}$ matrix, forming the scan context descriptor $\boldsymbol{SC} = \left(a_{i,j}\right) \in \mathbb{R}^{N_\mathrm{r} \times N_\mathrm{s}}$. Here, we encode bins separately with the maximum height and point density, preparing two descriptors, $\boldsymbol{SC}_\mathrm{h} = \left(h_{i,j}^{\max}\right) \in \mathbb{R}^{N_\mathrm{r} \times N_\mathrm{s}}$ and $\boldsymbol{SC}_\mathrm{d} = \left(d_{i,j}\right) \in \mathbb{R}^{N_\mathrm{r} \times N_\mathrm{s}}$. In Figure 5a, the color of each bin corresponds to the highest point within it, whereas the numerical value indicates the density of points contained therein.

#### 3.3.2. Co-Visible Region Detection

Scans from ground and aerial platforms differ in perspective, coverage, and resolution. The scans captured from the heterogeneous platforms have limited sensing range. Ground-based scans can cover most areas near the ground but information from higher places remains unknown. Conversely, aerial-based scans can capture information from higher elevations, but areas near the ground are susceptible to occlusion. These factors significantly increase the difficulty of directly matching the heterogeneous scans. Our approach involves detecting co-visible regions between heterogeneous scans and generating co-view context descriptors, as shown in Figure 5b.

To better illustrate the approach for detecting co-visible regions, a typical example of heterogeneous scans is presented in Figure 6. Figure 6a depicts a ground-based scan. Due to occlusion, all roof points, partial ground points, and partial facade points are not visible. The points captured from the ground-based scan are represented by green and yellow colors in Figure 6c. Figure 6b illustrates a general aerial-based scanning mode. The aircraft conducts a push-broom scan perpendicular to its forward direction, resulting in partial ground points, partial facade points, and all roof points being scanned. These points are represented by blue and green colors in Figure 6c. Based on this analysis, horizontal and vertical co-visible regions are detected separately.

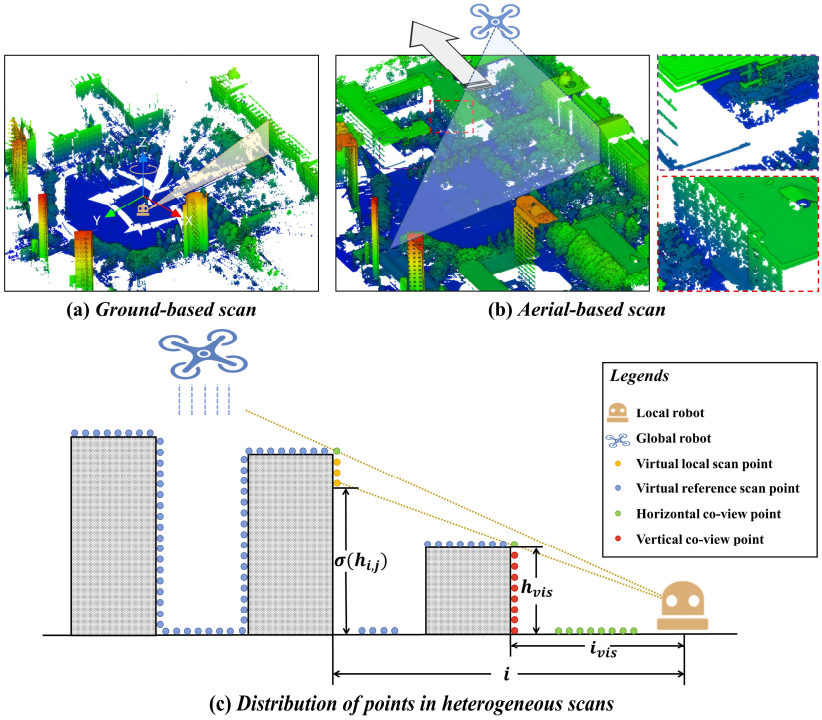


**Figure 6.** A typical example of ground-based and aerial-based scans.

The green points in Figure 6c are ground points and building outline points. They can be directly identified as the horizontal co-visible points. To find these points, a simple visibility analysis of the $\boldsymbol{SC}_{\mathrm{h}}$ descriptor is needed. Assuming the currently processing element is at the $i$-th row and $j$-th column of the descriptor matrix, denoted as $h_{i,j}^{\max}$, the horizontal visibility $\boldsymbol{V}_{\mathrm{hor}}\left(h_{i,j}^{\max}\right) \in \mathbb{R}^{N_{\mathrm{r}} \times N_{\mathrm{s}}}$ at the position $(i, j)$ is calculated as

$$\boldsymbol{V}_{\mathrm{hor}}\left(h_{i,\,j}^{\max}\right) = \begin{cases} \text{true}, & h_{i,\,j}^{\max} \geq \sigma_{\mathrm{h}}\left(h_{i,\,j}^{\max}\right) \\ \text{false}, & h_{i,\,j}^{\max} < \sigma_{\mathrm{h}}\left(h_{i,\,j}^{\max}\right) \end{cases} \tag{1}$$

$$\sigma_{\mathrm{h}}\left(h_{i,\,j}^{\max}\right) = i \times \frac{h_{\mathrm{vis}}}{i_{\mathrm{vis}}} \tag{2}$$

where $h_{\text{vis}}$ and $i_{\text{vis}}$ represent the maximum visible height and the corresponding row index in the previous moment, respectively. These two values will be updated if the current horizontal visibility is true.

The co-visible regions between aerial-based and ground-based scans include not only ground or building outline points but also some facade points. In Figure 6c, the red box indicates the scannable facades that are parallel to the flight direction, whereas the purple box indicates facades perpendicular to the flight direction, which cannot be scanned. In order to find the vertical points that are co-visible between aerial-based and ground-based scans, we use the $\boldsymbol{SC}_{\text{d}}$ descriptor to judge the facades and determine the vertical visibility $\boldsymbol{V}_{\text{ver}}(d_{i,j}) \in \mathbb{R}^{N_{\text{r}} \times N_{\text{s}}}$ at the position $(i, j)$ as

$$\boldsymbol{V}_{\text{ver}}(d_{i,j}) = \begin{cases} \text{true}, & d_{i,j} \geq \text{Avr}(\boldsymbol{SC}_{\text{d}}) \\ \text{false}, & d_{i,j} < \text{Avr}(\boldsymbol{SC}_{\text{d}}) \end{cases} \tag{3}$$

where $\text{Avr}(\boldsymbol{SC}_{\text{d}})$ represents the average point density of the $\boldsymbol{SC}_{\text{d}}$ descriptor. The visibility of each bin is indicated by "$\surd$" and "$\times$" in Figure 5b.

Based on the visibility analysis, we propose the co-view context descriptor $\boldsymbol{CoC} = (c_{i,j}) \in \mathbb{R}^{N_{\text{r}} \times N_{\text{s}}}$. For the bins that are only horizontally co-visible, we directly record their maximum height as $c_{i,j}$ in the co-view context descriptor. For example, in the sector shown in Figure 5a, the highest point within the innermost bin is blue, exhibiting true horizontal but false vertical visibility. Hence, it is still displayed as blue in the first row of the co-view context descriptor in Figure 5b. For the bins that are vertically co-visible, we compute the average height of the points within the bin as the value of $c_{i,j}$ to enhance the discrimination of the descriptor. For example, in the sector shown in Figure 5a, the bin of the second inner ring contains points of both red and blue colors, with true horizontal and vertical visibility. We display it as magenta in the second row of the co-view context descriptor in Figure 5b (we use magenta to indicate average since it comes from the mix of blue and red). As for the outermost bin with false horizontal visibility, we set it as black in the co-view context descriptor in Figure 5b, indicating that the value in this region is zero. The above descriptor generation approach can be formulated as

$$c_{i,j} = \begin{cases} h_{i,j}^{\max}, & \boldsymbol{V}_{\text{hor}}\left(h_{i,j}^{\max}\right) = \text{true} \ \wedge \ \boldsymbol{V}_{\text{ver}}(d_{i,j}) = \text{false} \\ h_{i,j}^{\text{avr}}, & \boldsymbol{V}_{\text{hor}}\left(h_{i,j}^{\max}\right) = \text{true} \ \wedge \ \boldsymbol{V}_{\text{ver}}(d_{i,j}) = \text{true} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $h_{i,j}^{\text{avr}}$ is the average height of the points within the bin at the position $(i, j)$.

### 3.3.3. Descriptor Matching

Figure 5c shows the matching process of the descriptor. In order to achieve rotational invariance, the co-view context descriptor is compressed into an $N_{\text{r}}$-dimensional vector as

$$\boldsymbol{k} = (\psi(r_i)) \in \mathbb{R}^{N_{\text{r}}}, \ i \in [1, N_{\text{r}}] \tag{5}$$

where $r$ represents each row of the co-view context descriptor and $\psi(\cdot)$ represents the occupancy ratio of a row vector using the $L_0$ norm:

$$\psi(r_i) = \frac{\| r_i \|_0}{N_{\text{s}}} \tag{6}$$

Due to the compression of the row vector, we can utilize a single value to describe the features of the horizontal 360°, thereby achieving rotational invariance of the descriptor with respect to the yaw angle.

In our approach, VRSs and VLSs are utilized as candidates and queries, respectively. Denoting the co-view context descriptors of the VRSs as $\{\boldsymbol{CoC}\}_M^{\text{R}}$ and those of the VLSs as

$\{\boldsymbol{CoC}\}_N^{\mathrm{L}}$, they are compressed to obtain vectors $\{\boldsymbol{k}\}_M^{\mathrm{R}}$ and $\{\boldsymbol{k}\}_N^{\mathrm{L}}$, where $M$ and $N$ represent the numbers of VRSs and VLSs, respectively. A KD tree is then constructed for $\{\boldsymbol{k}\}_M^{\mathrm{R}}$. For each querying $\boldsymbol{k}^{\mathrm{Q}} \in \{\boldsymbol{k}\}_N^{\mathrm{L}}$, the most similar vector $\boldsymbol{k}^{\mathrm{C}} \in \{\boldsymbol{k}\}_M^{\mathrm{R}}$ is retrieved. This process results in obtaining the most similar candidate co-view context descriptor $\boldsymbol{CoC}^{\mathrm{C}} \in \{\boldsymbol{CoC}\}_M^{\mathrm{R}}$ for each querying $\boldsymbol{CoC}^{\mathrm{Q}} \in \{\boldsymbol{CoC}\}_N^{\mathrm{L}}$. We adopt the cosine distance between two descriptors, $\boldsymbol{CoC}^{\mathrm{Q}}$ and $\boldsymbol{CoC}^{\mathrm{C}}$, as the distance metric, as follows:

$$\mathrm{d}\left(\boldsymbol{CoC}^{\mathrm{Q}}, \boldsymbol{CoC}^{\mathrm{C}}\right) = \frac{1}{N_{\mathrm{s}}} \sum_{j=1}^{N_{\mathrm{s}}} \left(1 - \frac{\boldsymbol{c}_j^{\mathrm{Q}} \cdot \boldsymbol{c}_j^{\mathrm{C}}}{\| \boldsymbol{c}_j^{\mathrm{Q}} \| \| \boldsymbol{c}_j^{\mathrm{C}} \|}\right) \tag{7}$$

where $\boldsymbol{c}_j^{\mathrm{Q}}$ and $\boldsymbol{c}_j^{\mathrm{C}}$ represent the $j$-th column vectors of the two descriptors, $\| \cdot \|$ denotes the $L_2$ norm, and the number of columns $N_{\mathrm{s}}$ is used for normalization.

Due to the different horizontal orientations between VRSs and VLSs, the column shift is also required for the candidate co-view context descriptors $\boldsymbol{CoC}^{\mathrm{C}}$ to obtain the best matched descriptor pair. Denoting $\boldsymbol{CoC}_n^{\mathrm{C}}$ the co-view context descriptor with its columns shifted $n$ units from the original $\boldsymbol{CoC}^{\mathrm{C}}$, we calculate the minimum distance between $\boldsymbol{CoC}^{\mathrm{Q}}$ and $\boldsymbol{CoC}^{\mathrm{C}}$ as

$$d_{\min} = \min_{n \in \{N_{\mathrm{s}}\}} \mathrm{d}\left(\boldsymbol{CoC}^{\mathrm{Q}}, \boldsymbol{CoC}_n^{\mathrm{C}}\right) \tag{8}$$

The number of the corresponding column shift is

$$n^* = \operatorname*{argmind}_{n \in \{N_{\mathrm{s}}\}} \left(\boldsymbol{CoC}^{\mathrm{Q}}, \boldsymbol{CoC}_n^{\mathrm{C}}\right) \tag{9}$$

### *3.4. Two-Phase Localization Strategy*

Our solution adopts a two-phase strategy for the global localization of heterogeneous scans, comprising a coarse step and a refined step.

As a mobile robot performs its scanning task, each $\boldsymbol{CoC}^{\mathrm{Q}}$ in $\{\boldsymbol{CoC}\}_N^{\mathrm{L}}$ will find the most similar $\boldsymbol{CoC}^{\mathrm{C}}$ in $\{\boldsymbol{CoC}\}_M^{\mathrm{R}}$, and the shortest distance constitutes the set $\{d_{\min}\}_N$. The minimum distance in $\{d_{\min}\}_N$ corresponds to the best matched descriptor pair between VLSs ang VRSs. In the co-view context descriptor, $N_{\mathrm{s}}$ is defined as the number of sectors. Each column shift therefore represents a yaw rotation of $2\pi/N_{\mathrm{s}}$. The rotation angle corresponding to $n^*$ column shifts is

$$yaw = \frac{2\pi n^*}{N_{\mathrm{s}}} \tag{10}$$

The coordinate transformation is performed on the virtual scan using the rotation angle $yaw$ obtained from descriptor matching and the coordinate translation $\boldsymbol{t}^{\mathrm{R}}$ and $\boldsymbol{t}^{\mathrm{L}}$. Let $\boldsymbol{P}^{\mathrm{R}}$ and $\boldsymbol{P}^{\mathrm{L}}$ denote the best matched VRS and VLS, $\boldsymbol{R} \in \mathrm{SO}(3)$ and $\boldsymbol{t} \in \mathbb{R}^3$ denote the rotation matrix and translation vector, and $\boldsymbol{T} = [\boldsymbol{R}|\boldsymbol{t}] \in \mathrm{SE}(3)$ denote the homogeneous transformation matrix. The scan after coarse localization is

$$\boldsymbol{P}_{\mathrm{coarse}}^{\mathrm{L}} = \boldsymbol{R}_{\mathrm{coarse}} \cdot \boldsymbol{P}^{\mathrm{L}} + \boldsymbol{t}_{\mathrm{coarse}} \tag{11}$$

where the rotation matrix $\boldsymbol{R}_{\mathrm{coarse}}$ and the translation vector $\boldsymbol{t}_{\mathrm{coarse}}$ are

$$\begin{aligned} \boldsymbol{R}_{\mathrm{coarse}} &= \begin{bmatrix} \cos(yaw) & -\sin(yaw) & 0 \\ \sin(yaw) & \cos(yaw) & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \boldsymbol{t}_{\mathrm{coarse}} &= \boldsymbol{t}^{\mathrm{R}} - \boldsymbol{t}^{\mathrm{L}} = \left[x^{\mathrm{R}} - x^{\mathrm{L}}, y^{\mathrm{R}} - y^{\mathrm{L}}, z^{\mathrm{R}} - z^{\mathrm{L}}\right]^{\mathbf{T}} \end{aligned} \tag{12}$$

Here, $\boldsymbol{t}^{\mathrm{r}}$ and $\boldsymbol{t}^{\mathrm{l}}$ are derived from the construction processes of VRSs and VLSs, respectively. The coarse transformation is $\boldsymbol{T}_{\mathrm{coarse}} = [\boldsymbol{R}_{\mathrm{coarse}}|\boldsymbol{t}_{\mathrm{coarse}}]$.

By checking the vertical visibility of the matched VRS and VLS descriptors, points that exhibit significant differences in vertical visibility are identified and eliminated. This process aims to reduce the impact of inconsistent points, primarily those facade points not scanned by the VRS, on the refined localization step. Taking the coarse localization as the initial value, trimmed ICP [20] is employed to perform fine registration between point clouds of the VRS and VLS. The trimmed ICP algorithm leverages the least trimmed squares (LTS) method to fit the error function

$$\mathrm{E}(\boldsymbol{R}_{\text{refined}}, \boldsymbol{t}_{\text{refined}}) = \frac{1}{m}\sum_{k=1}^{m}\left\|\boldsymbol{P}^{\mathrm{R},\,k} - \left(\boldsymbol{R}_{\text{refined}}\cdot\boldsymbol{P}_{\text{coarse}}^{\mathrm{L},\,k} + \boldsymbol{t}_{\text{refined}}\right)\right\| \tag{13}$$

where $m$ represents the number of correspondences, and $\boldsymbol{P}^{\mathrm{R},\,k}$ and $\boldsymbol{P}_{\text{coarse}}^{\mathrm{L},\,k}$ represent a set of corresponding points in the VRS point cloud and the coarse-localized VLS point cloud, respectively. The LTS method sorts the residuals calculated from each set of corresponding points in ascending order and retains only the top fraction $\epsilon$ to fit the error function. This algorithm can effectively remove outlier correspondences resulting from the low overlap between heterogeneous scans. The refined transformation $\boldsymbol{T}_{\text{refined}} = [\boldsymbol{R}_{\text{refined}}|\boldsymbol{t}_{\text{refined}}]$ is determined by iteratively minimizing the error function. The final transformation matrix from the local scan to the reference scan after the two-phase localization is given by

$$\boldsymbol{T}_{\mathrm{L-R}} = \boldsymbol{T}_{\text{refined}}\cdot\boldsymbol{T}_{\text{coarse}} \tag{14}$$

We simulate a scenario in which robots in a multi-robot system are unable to communicate with each other. Each robot calculates its transformation matrix with respect to the pre-set base map by localizing its own scan in the reference scan. This process enables global localization of heterogeneous multi-robot systems in scenarios where communication is impaired.

## 4. Experiments

We implement the proposed solution in C++ on a Linux Ubuntu system and test it on an industrial computer equipped with 8 GB RAM and an ARM Cortex-A55@1.8GHz CPU. This low configuration is a fair representation of the mobile robot performance.

### *4.1. Datasets Description*

We evaluate the performance of the proposed global localization solution using three datasets collected from different scenarios, namely construction, gymnasium, and campus. These datasets were captured at Wuhan University, respectively around the Forepart Construction, inside and outside the Zhuoer Gymnasium, and around the Friendship Square, as depicted in Figure 7. They were sourced from four distinct heterogeneous platforms: unmanned aerial vehicle (UAV), terrestrial station, trolley, and handheld. Candidate co-view context descriptors generated by the pre-set base map were assigned to multi-robot systems to conduct local scanning for online or offline global localization purposes. The detailed information about the datasets can be found in Table 1.


(a) Forepart Construction (b) Zhuoer Gymnasium (c) Friendship Square

**Figure 7.** Study areas for the three datasets.

**Table 1.** Details of the datasets.

| Dataset | Base Map | | Local Scan | | | | | Average Overlap |
|---|---|---|---|---|---|---|---|---|
| | Scanner | Density (Points/m$^2$) | Index | Data Source | Scanner | Density (Points/m$^2$) | Ground Truth | |
| Construction | LJYY-FT1500 | 88.4 | 1 | WHU-TLS Heritage building 3 | VZ-400 | 3348.3 | Provided by WHU-TLS | 57.8% |
| | | | 2 | WHU-TLS Heritage building 4 | VZ-400 | 5536.4 | Provided by WHU-TLS | 50.6% |
| | | | 3 | WHU-TLS Heritage building 5 | VZ-400 | 3412.9 | Provided by WHU-TLS | 36.4% |
| | | | 4 | WHU-TLS Heritage building 6 | VZ-400 | 5876.5 | Provided by WHU-TLS | 52.1% |
| Gymnasium | CHCNAV-AU20 | 261.3 | 1 | Gymnasium Self-built 1 | Mid-360 | 429.8 | Handcraft + T-ICP | 49.9% |
| | | | 2 | Gymnasium Self-built 2 | Mid-360 | 913.0 | Handcraft + T-ICP | 4.7% |
| | | | 3 | Gymnasium Self-built 3 | Mid-360 | 252.4 | Handcraft + T-ICP | 22.6% |
| Square | LJYY-FT1500 | 88.2 | 1 | WHU-TLS Campus 1 | VZ-400 | 3440.0 | Provided by WHU-TLS | 74.8% |
| | | | 2 | WHU-TLS Campus 2 | VZ-400 | 1782.2 | Provided by WHU-TLS | 65.0% |
| | | | 3 | WHU-TLS Campus 3 | VZ-400 | 2255.0 | Provided by WHU-TLS | 84.4% |
| | | | 4 | WHU-TLS Campus 4 | VZ-400 | 2700.8 | Provided by WHU-TLS | 85.1% |
| | | | 5 | Square Self-built 1 | Mid-360 | 90.8 | Handcraft + T-ICP | 54.1% |
| | | | 6 | Square Self-built 2 | RS-Bpearl | 824.1 | Handcraft + T-ICP | 43.7% |

LJYY-FT1500 [52] is an airborne laser radar system developed by LuoJiaYiYun. It was mounted on a multi-rotor DJI M300 RTK [53] UAV to generate base maps for the construction and campus datasets. During the operation, the UAV flew automatically on a predefined course at an altitude of approximately 300 m and a speed of about 8 m/s. CHCNAV-AU20 [54] is a multi-platform-available mobile mapping system equipped with an integrated navigation system (INS) to obtain accurate laser scans. We fixed it on a mobile trolley to provide the base map for the gymnasium dataset. It is worth noting that the data collection process for base maps follows general surveying standards and has not been deliberately densified or refined. The 3D terrestrial laser scanner RIEGL VZ-400 [55] is capable of providing long-range and high-precision point clouds. In the public benchmark WHU-TLS [56], the WHU-TLS heritage building dataset and the WHU-TLS campus dataset were captured using the RIEGL VZ-400 laser scanner, which we consider as a local scan from a stationary robot. The self-built data come from LiDAR SLAM. Livox Mid-360 [57] is a mixed solid-state LiDAR that delivers 3D perception in 360 degrees. We used a handheld device and employed the FAST-LIO2 algorithm [58] for localization, treating each local point cloud as provided by a single robot. Robosense bpearl is a super-wide field-of-view (FOV) mechanical LiDAR with 32 hemispherical scan lines. We utilized the handheld device prototype designed by Duan [59] as a single robot to provide a local scan of the campus dataset. The scan points are downsampled since our bin-wise descriptor does not require dense point clouds. The overlaps between the base map and local scans range from 4.7% to 85.1%, and the point densities vary significantly. These differences pose significant challenges for global localization. The laser scanners and platforms used in the experiments are shown in Figure 8.


**Figure 8.** Laser scanners and platforms used in the experiments.

### *4.2. Evaluation Criteria*

The global localization accuracy is evaluated in terms of the axis-angle rotation error, translation error, and successful localization rate (SLR). Letting the estimated transformation matrix from the local scan to base map be $[\boldsymbol{R}_{\mathrm{L-R}}|\boldsymbol{t}_{\mathrm{L-R}}]$ and the ground truth be $[\boldsymbol{R}_{\mathrm{GT}}|\boldsymbol{t}_{\mathrm{GT}}]$, the axis-angle rotation error $e_\theta$ and translation error $e_\mathrm{t}$ are calculated as

$$\begin{aligned} e_\theta &= \operatorname{across}\left(\frac{\operatorname{tr}\left(\boldsymbol{R}_{\mathrm{L-R}} \cdot \boldsymbol{R}_{\mathrm{GT}}{}^{-1}\right) - 1}{2}\right) \\ e_\mathrm{t} &= \parallel \boldsymbol{t}_{\mathrm{L-R}} - \boldsymbol{t}_{\mathrm{GT}} \parallel \end{aligned} \tag{15}$$

where $\mathrm{TR}(\cdot)$ denotes the trace of a matrix.

We use the successful localization (SL) to indicate whether a global localization attempt is successful. SL is defined as

$$\mathrm{SL} = \begin{cases} 1, & e_\theta < \sigma_\theta \wedge e_\mathrm{t} < \sigma_\mathrm{t} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $\sigma_\theta$ and $\sigma_t$ are predefined thresholds for rotation and translation errors, respectively. In this paper, they are set to 5.0° and 3.0 m, respectively. The SLR is then calculated as

$$\mathrm{SLR} = \frac{N_s}{N_{\mathrm{total}}} \tag{17}$$

where $N_s$ is the number of successful localizations and $N_{\mathrm{total}}$ is the total number of localization attempts.

### *4.3. Parameter Settings*

Table 2 shows the parameter settings of the proposed solution. The subsequent experiments are conducted based on these parameter settings.

**Table 2.** The parameter settings of the proposed solution.

| Parameters | Symbol | Description | Value |
|---|---|---|---|
| Block size | $L$ | The block size of reference scan | 2.0 m |
| Search radius | $r_{ref}$ | The search radius of reference scan | 50 m for LiDAR SLAM and 100 m for TLS |
| Distance threshold | $\sigma_\mathrm{d}$ | The distance threshold for keyframe selection | 2.0 m |
| Radial partition | $N_\mathrm{r}$ | The number of radial partitions | 20 |
| Azimuthal partition | $N_\mathrm{s}$ | The number of azimuthal partitions | 60 |
| Descriptor range | $r_{max}$ | The max range of co-view context descriptor | 50 m for LiDAR SLAM and 100 m for TLS |
| Overlap rate | $\epsilon$ | The overlap rate for trimmed ICP | 60% |

The parameters $L$ and $\sigma_\mathrm{d}$ are the block size of reference scan and the distance threshold for keyframe selection, respectively, which impact the number of VRSs and VLSs. Smaller $L$ and $\sigma_\mathrm{d}$ result in more virtual scans with shorter intervals, leading to better localization effectiveness but longer runtime and vice versa. Typically, $L$ and $\sigma_\mathrm{d}$ are set to equal to maintain a consistent sampling interval. Figure 9 illustrates the SLR and average runtime for each dataset under different configurations of $L$ and $\sigma_\mathrm{d}$. When $L$ and $\sigma_\mathrm{d}$ are set to 1.0 m or 2.0 m, all three datasets maintain a 100% SLR. As $L$ and $\sigma_\mathrm{d}$ increase further, the SLR gradually decreases. Balancing the runtime efficiency and localization effectiveness, these two parameters are set to 2.0 m in our experiments.

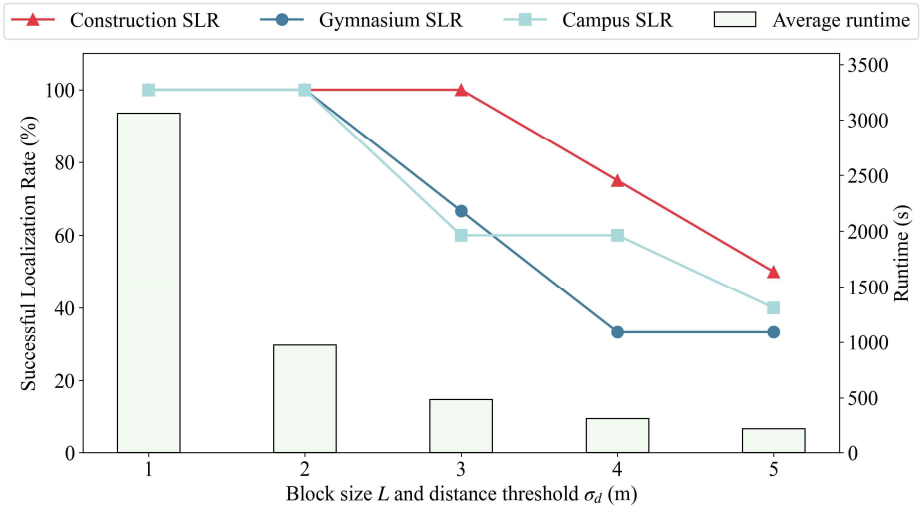


**Figure 9.** Parameter tests for VRS block size and VLS keyframe distance threshold.

The parameter $r_{\mathrm{ref}}$ is the search radius for constructing VRSs. It is adjusted according to the measurement range of the laser scanner. If the local scan is obtained by an MLS method like LiDAR SLAM, it is initially processed to construct VLSs, which are then matched with the VRSs. In this case, the search radius $r_{\mathrm{ref}}$ is set to the measurement range

of the mobile laser scanning device, typically around 50 m. If the local scan originates from single-station laser scanning, such as TLS, the entire scan is matched with the VRS. Therefore, the parameter $r_{ref}$ is set larger, typically around 100 m, corresponding to the measurement range of TLS devices.

The parameters $N_r$, $N_s$, and $r_{max}$ are associated with the co-view context descriptor. $N_r$ and $N_s$ control the number of bins in the descriptor. They are set to 20 and 60 by default based on the experimental findings from scan context [6] and skyline context [34]. The parameter $r_{max}$ controls the range of the descriptor and is typically set to be equal to $r_{ref}$.

The parameter $\epsilon$ represents the overlap rate between point clouds in trimmed ICP. We empirically set it to 60%.

### 4.4. Experiment Results

Table 3 lists the number of VLSs and VRSs generated from each dataset during the global localization process. Local scans from the WHU-TLS dataset have only one VLS because we consider them as stationary robots. Figures 10–12 illustrate the results of global localization for the three datasets and offer additional details. The point clouds in the figures are displayed from a bird's-eye view. In the overview, the base maps are depicted by gray points, while the local scans are distinguished by different colors. In the detailed views, the local scans are visualized with colored ribbons to indicate elevation. The localization results depicted in the figures demonstrate that our solution effectively addresses various scenarios involving heterogeneous scans and achieves good performance.

**Table 3.** The number of VRSs and VLSs.

| Reference Scan | #VRS | Local Scan | #VLS |
|---|---|---|---|
| Construction ALS point cloud | 1535 | WHU-TLS Heritage building 3 | 1 |
| | | WHU-TLS Heritage building 4 | 1 |
| | | WHU-TLS Heritage building 5 | 1 |
| | | WHU-TLS Heritage building 6 | 1 |
| Gymnasium MLS point cloud | 9527 | Gymnasium Self-built 1 | 223 |
| | | Gymnasium Self-built 2 | 245 |
| | | Gymnasium Self-built 3 | 176 |
| Square ALS point cloud | 8360 | WHU-TLS Campus 1 | 1 |
| | | WHU-TLS Campus 2 | 1 |
| | | WHU-TLS Campus 3 | 1 |
| | | WHU-TLS Campus 4 | 1 |
| | | Square Self-built 1 | 154 |
| | | Square Self-built 2 | 719 |

In the construction dataset, the base map is provided by ALS, while all local scans are acquired by TLS. This represents a special case in multi-robot global localization where all local robots are stationary. In such a scenario, the step of VLS construction can be omitted, and the global localization problem simplifies into a pure LiDAR place recognition and point cloud registration problem. However, the differences in scanning perspectives, ranges, and resolutions between ALS and TLS significantly increase the difficulty of registration. In Figure 10, all four local scans are successfully localized in the base map, demonstrating that the co-view context descriptor proposed in this paper can cope well with the heterogeneous scans.

In the gymnasium dataset, the reference and local scans are obtained homogeneously using the MLS method. This exemplifies a typical case in multi-robot global localization where all local robots are in motion. In this dataset, the scanning scenarios encompass both indoor and outdoor environments with limited overlaps, and the three local robots did not encounter each other during the operation. With the support of the proposed communication-free solution, each robot can perform local and global localization simulta-

neously under the guidance of the pre-set base map. As shown in Figure 11, the successful localization indicates that our base-map-guided global LiDAR localization solution is effective.

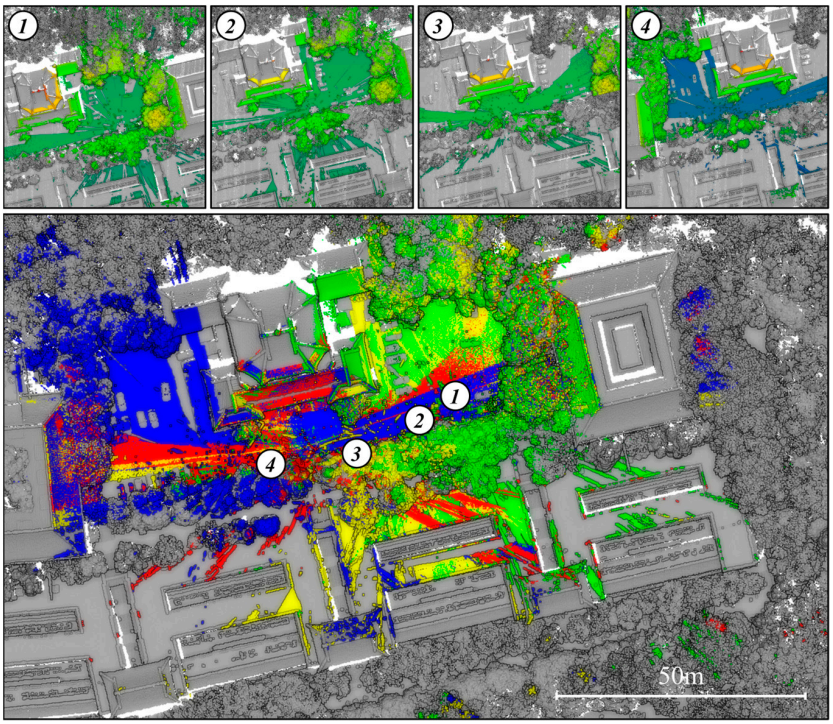


**Figure 10.** The localization results of the construction dataset. The **bottom** figure shows an overview of the localization results, and the **top** figures show detailed views of each localized scan.

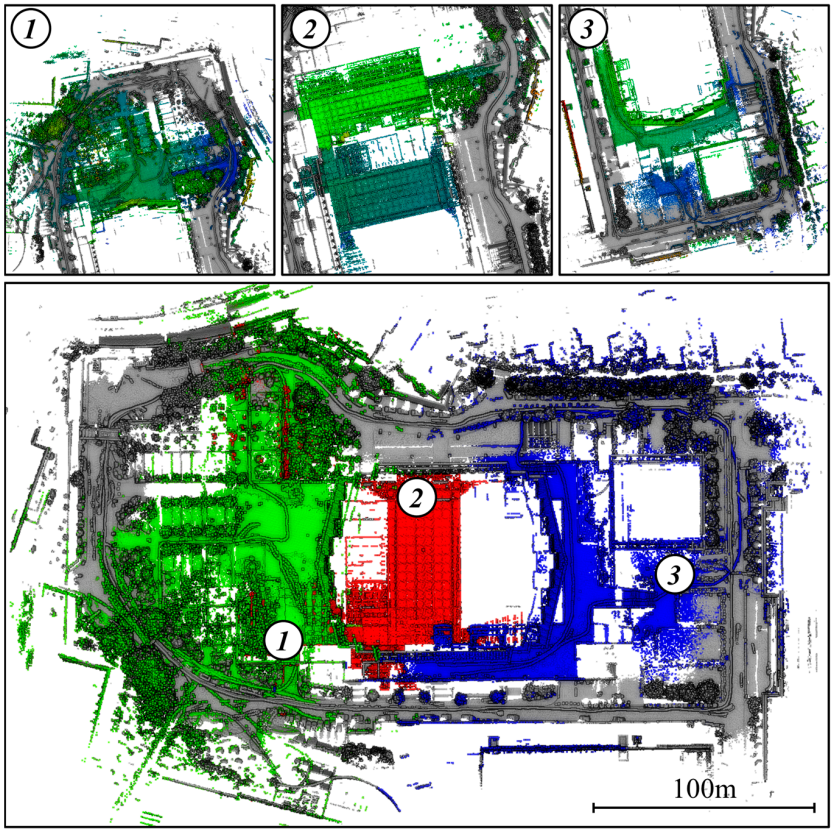


**Figure 11.** The localization results of the gymnasium dataset. The **bottom** figure shows an overview of the localization results, and the **top** figures show detailed views of each localized scan.

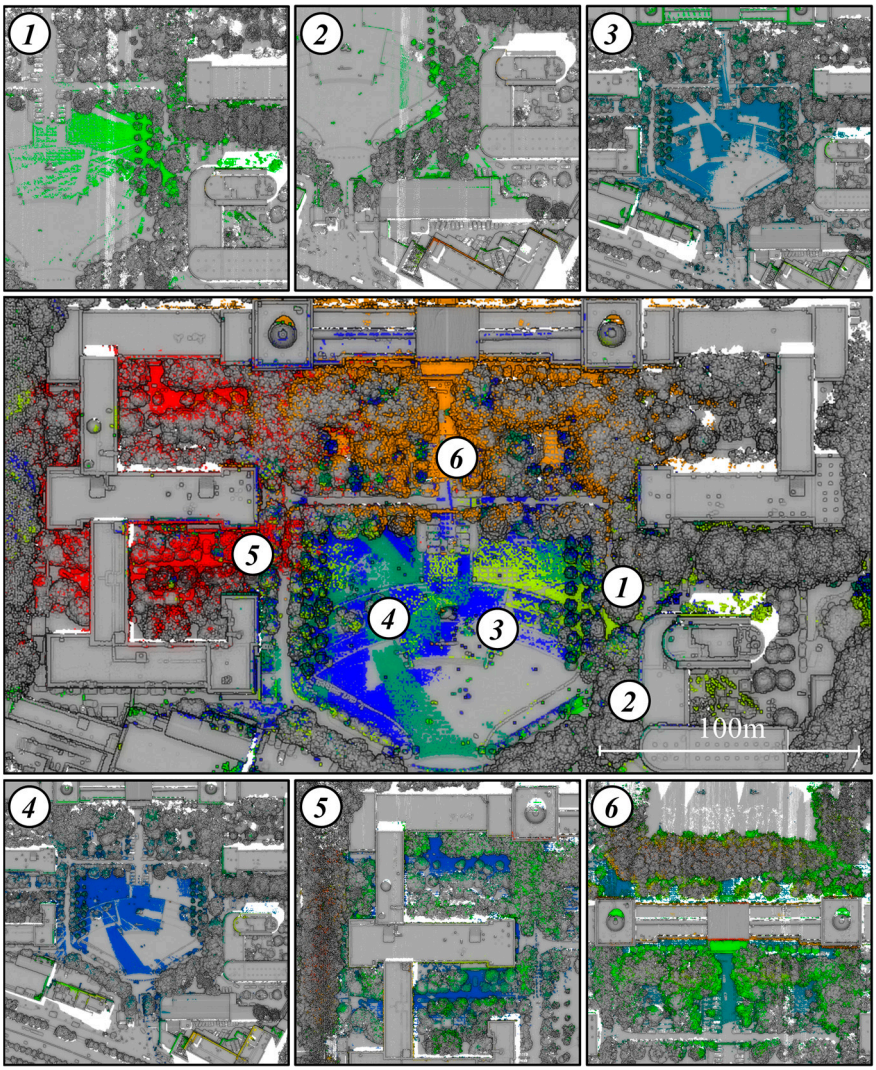


**Figure 12.** The localization results of the campus dataset. The **middle** figure shows an overview of the localization results, and the **top** and **bottom** figures show detailed views of each localized scan.

In the square dataset, the base map is provided by ALS, while the local scans consist of TLS and MLS acquired using different devices. We consider this to be a complex scenario in multi-robot systems global localization, as the scans are totally hybrid. The dataset comprises four different laser scanning modes: aerial-based laser scanning, ground-based single-station laser scanning, mechanical LiDAR mobile scanning, and mixed solid-state LiDAR mobile scanning. The variations arise not only between the local scan and the base map but also among local scans themselves, posing huge challenges for heterogeneous multi-robot systems global localization. However, due to the effective extraction of co-visible regions by the proposed discriminative co-view context descriptor, all six sets of heterogeneous scans in Figure 12 can be successfully localized in the base map.

### *4.5. Quantitative Analysis*

#### 4.5.1. Successful Localization Rate Validation

To validate the successful localization rate of our approach, a total of 30 evenly distributed local scans were additionally collected using a handheld Livox Mid-360 [57] laser scanning device on the campus of Wuhan University. These 30 local scans are globally localized with guidance from the ALS base map, as shown in Figure 13. The SLR is calculated according to Equation (17). Among these, 28 local scans are successfully globally localized, resulting in a successful localization rate of 93.3%.

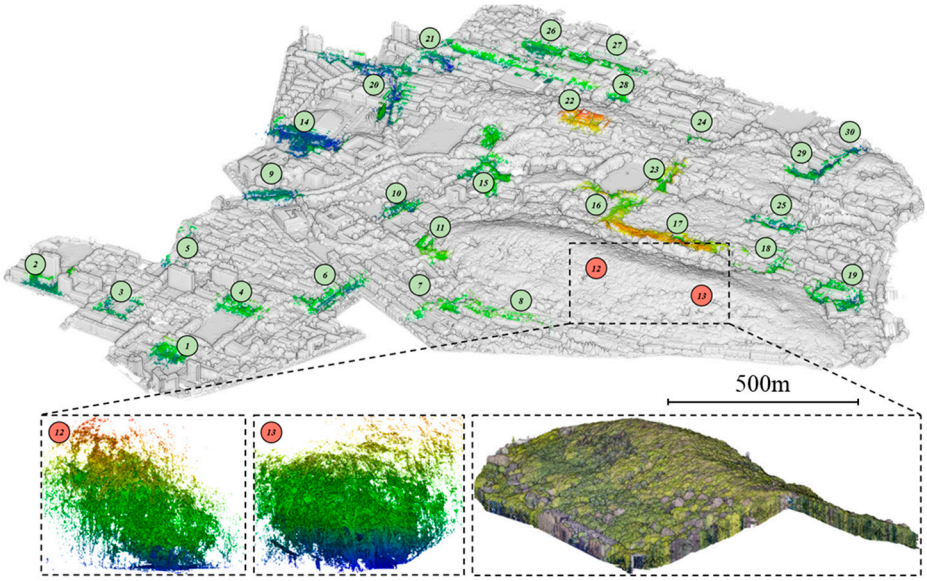


**Figure 13.** The localization results of the SLR validation experiment. Green labels indicate successfully localized scans, while red labels indicate failed localized scans.

The 12th and 13th local scans are not successfully localized. The reason is that these two local scans were collected on Luojia Mountain, where the trees are tall and dense. On the one hand, the dense canopy almost completely obscures the ground in the ALS base map, preventing the construction of virtual scans. On the other hand, the lack of distinguishable features among the trees leads to ambiguities during the registration of local scans with the base map.

#### 4.5.2. Accuracy Evaluation

According to Equation (15), the rotation and translation errors of the proposed localization approach are evaluated, as listed in Table 4. In the public benchmark WHU-TLS, ground truth transformations between neighboring point clouds are provided. We calculate the transformation between two neighboring point clouds after global localization ($\boldsymbol{R}_{\mathrm{L-R}}$ and $\boldsymbol{t}_{\mathrm{L-R}}$ in Equation (15)), and it is compared with the ground truth ($\boldsymbol{R}_{\mathrm{GT}}$ and $\boldsymbol{t}_{\mathrm{GT}}$ in Equation (15)) to reflect the localization accuracy. As for the self-built datasets, we roughly localize the local scan to the base map in a handcraft manner and then compute the ground truth using the trimmed ICP method. The transformation between the local scans and the base map ($\boldsymbol{R}_{\mathrm{L-R}}$ and $\boldsymbol{t}_{\mathrm{L-R}}$ in Equation (15)) is calculated and then compared with the ground truth ($\boldsymbol{R}_{\mathrm{GT}}$ and $\boldsymbol{t}_{\mathrm{GT}}$ in Equation (15)) to show the localization accuracy.

It is found that the rotation and translation errors range from 0.013° to 1.391° and from 0.058 m to 0.675 m, respectively. The rotation error between WHU-TLS heritage building 4 and WHU-TLS heritage building 5 stands out as significantly larger than others. This result stems from the symmetrical structure of the constructions, which causes the descriptors to be ambiguous in estimating the yaw angle. Moreover, the rotation and translation errors between the Square Self-built 2 and Square ALS point cloud are both large, with two underlying reasons. First, limited by sensor performance and SLAM algorithms, the local scan accuracy of Square Self-built 2 is inherently lower compared to scans captured by other scanning modes. In Figure 14, it is evident that the walls in Square Self-built 2 exhibit layering due to repeated scanning, which is not observed in the Square ALS point cloud. Second, the region of Square Self-built 2 is densely covered with trees that obscure the co-visibility between the aerial-based scan and the ground-based scan from different perspectives. These two factors together account for the significant error.

**Table 4.** The rotation and translation errors of each scan.

| Dataset | To Be Aligned | Reference | Rotation Error (deg) | Translation Error (m) |
|---|---|---|---|---|
| Construction | WHU-TLS Heritage building 3 | WHU-TLS Heritage building 4 | 0.091 | 0.058 |
| | WHU-TLS Heritage building 4 | WHU-TLS Heritage building 5 | 1.391 | 0.185 |
| | WHU-TLS Heritage building 5 | WHU-TLS Heritage building 6 | 0.103 | 0.071 |
| | WHU-TLS Heritage building 6 | Construction ALS point cloud | 1.065 | 0.065 |
| Gymnasium | Gymnasium Self-built 1 | Gymnasium MLS point cloud | 0.013 | 0.080 |
| | Gymnasium Self-built 2 | Gymnasium MLS point cloud | 0.045 | 0.120 |
| | Gymnasium Self-built 3 | Gymnasium MLS point cloud | 0.097 | 0.184 |
| Square | WHU-TLS Campus 1 | WHU-TLS Campus 2 | 0.018 | 0.256 |
| | WHU-TLS Campus 2 | WHU-TLS Campus 3 | 0.080 | 0.069 |
| | WHU-TLS Campus 3 | WHU-TLS Campus 4 | 0.017 | 0.145 |
| | WHU-TLS Campus 4 | Square ALS point cloud | 0.054 | 0.453 |
| | Square Self-built 1 | Square ALS point cloud | 0.114 | 0.169 |
| | Square Self-built 2 | Square ALS point cloud | 0.067 | 0.675 |

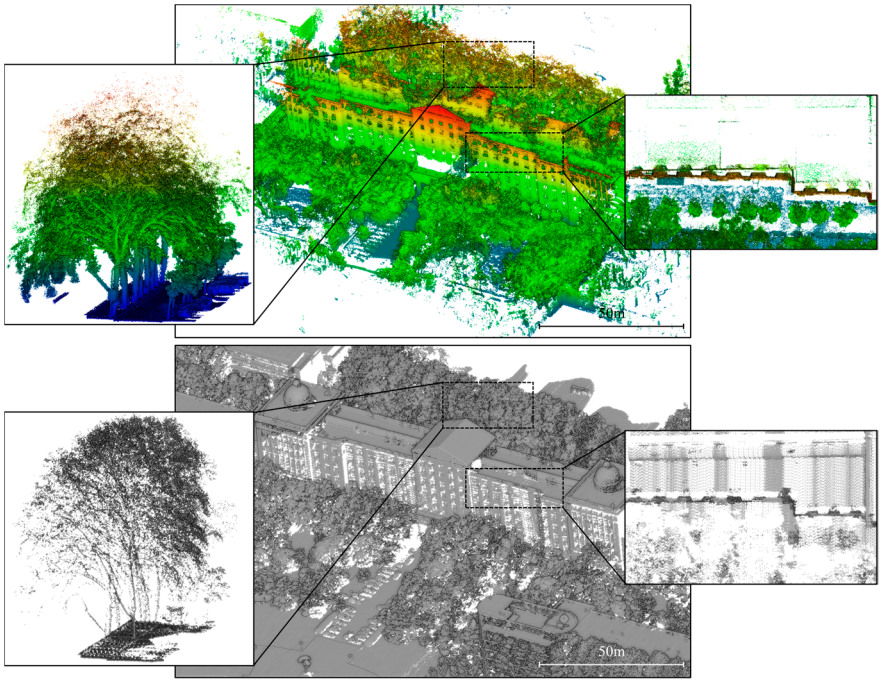


**Figure 14.** Comparison between Campus Self-built 2 and Campus ALS point cloud. The **top** figure shows the Campus Self-built 2 scan and its details. The **bottom** figure shows the Campus point cloud scan and its details.

We compare the performance of the proposed method with two state-of-the-art multi-view point cloud registration methods: the geometry-based HL-MRF method [27] and the deep-learning-based SGHR method [31]. The following work is done to ensure the fairness of the comparison: first, in each set of experiments, the base map is used to provide a reference for registration. Second, the point clouds transformed by the SGHR method are further refined with trimmed ICP alignment, as this refinement step is not included in the end-to-end process of SGHR. Table 5 lists the quantitative evaluation results of the global localization. It is important to note that we only included scans that meet the successful localization criteria in the calculation of the mean error and root mean square error (RMSE). Otherwise, the localization errors from failed attempts would significantly skew the results.

**Table 5.** Quantitative evaluation of global localization.

| Dataset | Method | Rotation Error (deg) | | Translation Error (m) | | SLR (%) |
|---|---|---|---|---|---|---|
| | | Average | RMSE | Average | RMSE | |
| Construction | HL-MRF | **0.056** | **0.058** | **0.012** | **0.014** | 75% |
| | SGHR | 0.337 | 0.337 | 0.302 | 0.302 | 25% |
| | Ours | 0.663 | 0.879 | 0.095 | 0.108 | **100%** |
| Gymnasium | HL-MRF | 0.315 | 0.346 | 0.661 | 0.724 | 100% |
| | SGHR | 1.697 | 1.812 | 3.186 | 3.593 | 75% |
| | Ours | **0.052** | **0.062** | **0.128** | **0.135** | **100%** |
| Square | HL-MRF | 0.096 | 0.102 | **0.109** | **0.155** | 50% |
| | SGHR | 0.815 | 1.031 | 0.996 | 1.180 | 33% |
| | Ours | **0.058** | **0.068** | 0.295 | 0.361 | **100%** |

Our method maintains an advantage in the SLR criteria, achieving a 100% successful localization rate for all three datasets, with average rotation and translation errors within 1° and 0.30 m, respectively. For the construction dataset, the HL-MRF method exhibits the highest rotation and translation accuracy. This is because the four local scans in the construction dataset all originate from WHU-TLS dataset. The HL-MRF method, specifically designed for multi-view point cloud registration, can accurately solve the relative poses of the homogeneous local scans. However, a 75% SLR of HL-MRF means that one out of four localizations failed. This failure occurred between the WHU-TLS heritage building 6 and the construction ALS point cloud. This indicates that while the HL-MRF method can effectively perform mutual localization of homogeneous scans, it struggles with global localization of heterogeneous scans across different viewpoints due to a lack of targeted design. In the gymnasium dataset, the low overlap between local scans and the reference scan results in larger errors for both the HL-MRF and SGHR methods. Our method constructs virtual scans based on the constraints of the robot's trajectory. This approach can be viewed as generating multiple smaller data processing units for place recognition. Consequently, it ensures that our method can effectively handle low-overlap cases. The square dataset consists of scans obtained through four different scanning modes, leading to significant variations in scanning perspectives, ranges, resolutions, and even periods among local scans and the base map. The heterogeneity poses a huge challenge in establishing associations between scans using the HL-MRF and SGHR methods. Nevertheless, our method still maintains sub-meter-level global localization accuracy for this complex dataset.

#### 4.5.3. Runtime Analysis

The time consumption during the local scanning process of the gymnasium dataset is listed in Table 6. Since the candidate descriptors are pre-generated and stored in each robot, we solely focus on the time consumed by a single robot for generating descriptors, detecting co-visible regions, and matching descriptors during online operation. It is found that the runtime is primarily concentrated on descriptor matching as it involves similarity calculation with candidate descriptors. For each laser frame, the total runtime is about 50 ms. Considering that the typical scanning frequency of LiDAR used for SLAM is around 10 Hz, we conclude that the proposed approach based on co-view context descriptors can adequately meet the requirements for online processing.

**Table 6.** Runtime analysis of the gymnasium dataset (ms).

| Sequence | Descriptor Generation | Co-Visible Region Detection | Descriptor Matching | Total |
|---|---|---|---|---|
| Self-built 1 | 1.522 | 1.973 | 47.715 | 51.210 |
| Self-built 2 | 2.723 | 4.227 | 47.202 | 54.152 |
| Self-built 3 | 1.490 | 1.979 | 47.198 | 50.667 |
| Average | 1.912 | 2.726 | 47.372 | 52.010 |

## 5. Discussion

In general, researchers aim to achieve global localization for multiple robots without relying on prior information, such as pre-set base maps. This is considered a standard collaborative SLAM problem. However, in specific scenarios—such as in mapping tasks where high-precision point cloud maps are required, in complex environments where mutual overlapping scans are difficult to obtain, or in applications where data fusion between heterogeneous platforms is involved—the base-map-guided LiDAR localization solution for heterogeneous robots proposed in this paper demonstrates superior performance. From a strategic perspective, we address the LiDAR global localization problem in multi-robot systems as a multi-view point cloud registration issue that leverages the prior knowledge from pre-set base maps. This approach naturally achieves superior results compared to existing methods, where multi-robot LiDAR localization without prior knowledge depends exclusively on overlapping regions between robots to estimate their global poses. For example, in a chain arrangement of three robots, the second robot shares overlapping regions with both the first and third robots, while the first and third robots do not overlap with each other. Using the coordinate system of the first robot as the global reference, the global pose of the second robot can be determined based on its overlap with the first robot, followed by the derivation of the third robot's global pose based on its overlap with the second one. However, this process results in cumulative localization errors from the second to the third robot, and as the number of robots increases, this accumulation further undermines the reliability of subsequent robots' localization. In contrast, our approach utilizes the base map as a global reference, directly aligning each robot to the base map's coordinate system, thereby effectively preventing error accumulation. Consequently, the global localization strategy proposed in this study outperforms current state-of-the-art geometry-based [27] and deep-learning-based [31] methods in both localization success rate and accuracy.

Nevertheless, localizing heterogeneous point clouds remains challenging, and addressing this issue is a central focus of our work., localizing heterogeneous point clouds remains challenging, and addressing this issue is a central focus of our work. In this paper, we develop a novel co-view context descriptor capable of detecting co-visible regions in heterogeneous point clouds. Inspired by the scan context descriptor [6], the co-view context descriptor enables efficient and concise feature encoding of point clouds. In addition to this, it provides adaptability to heterogeneous point clouds—an aspect generally overlooked by conventional registration methods. The primary challenges in heterogeneous aerial and ground point cloud localization involve the differences in scanning viewpoint, perspectives, ranges, and resolutions. Our approach begins by constructing virtual scans and uses ground information to mitigate viewpoint differences between the aerial and ground point clouds. Additionally, the range of the virtual scan is adjusted according to the sensor's configuration to eliminate the differences in scanning ranges. A descriptor is then constructed using the bin-wise encoding method, in which a representative feature is selected for all points within each bin, thus mitigating the impact of resolution differences. Finally, co-visible regions between aerial and ground platforms are identified at the descriptor level, effectively addressing the influence of perspectives differences. These advantages are clearly demonstrated in comparative experiments, where our method exhibits superior performance when the base map and robot point clouds are derived from scans taken from different sources. Moreover, we believe that the co-view context descriptor holds great potential and can be applied to various other scenarios, such as real-time data association in heterogeneous aerial–ground collaborative SLAM.

However, the proposed method exhibits limitations when processing regions densely populated with trees, both in terms of the successful localization rate and localization accuracy. Two main factors likely contribute to this: first, from a ground-level perspective, different trees often exhibit similar structural features, making it challenging to achieve high localization accuracy for individual robots; second, the dense tree canopy obstructs the line of sight between aerial and ground platforms, limiting the effectiveness of the co-view

context descriptor and thereby impacting multi-robot localization under the guidance of the base map. Future work will need to incorporate data fusion with other sensors to address the challenge of achieving robust localization in forested environments.

## 6. Conclusions

In this paper, we propose a global LiDAR localization solution for heterogeneous robots using a pre-set base map as a reference. This solution can be regarded as conducting place recognition of small-scale point clouds within a large-scale point cloud or as globally registering point clouds acquired through different laser scanning modes. By independently localizing each scan under the guidance of the pre-set base map, the risk of single-point failures or communication disruptions in global localization can be mitigated. Experimental results demonstrate that the proposed solution achieves good localization performance across different environments, robot types, scanning modes, and point cloud characteristics, offering reliable localization support for mobile robots. After global localization, the fused point cloud can compensate for blind spots and scanning defects in different perspectives, thereby enhancing the ability to represent indoor and outdoor scenes uniformly. In conclusion, our solution offers great application scenarios for existing HD maps. It can effectively support various downstream tasks such as geospatial information extraction, cultural heritage preservation, and real-world 3D reconstruction.

Future work will focus on two aspects. First, the proposed approach only estimates the pose relationship between the local scan and the base map. Considering scenarios where there are overlapping regions between multiple local scans, a multi-view point cloud registration step will be introduced to establish constraints between mobile robots. Further pose optimization will be conducted to achieve global consistency. Second, we currently only consider LiDAR as the external sensor for global localization. The co-view context descriptor will be expanded to adapt various sensors (e.g., RGB cameras, millimeter-wave radar, etc.) to address global localization problems for cross-modal heterogeneous mobile robots.

**Author Contributions:** Conceptualization, P.Z.; data curation, C.X.; formal analysis, P.Z.; funding acquisition, P.Z. and Q.H.; investigation, X.D.; methodology, X.D.; project administration, M.W.; resources, M.W.; software, C.X.; supervision, Q.H.; validation, M.W. and C.X.; visualization, X.D.; writing—original draft, X.D.; writing—review and editing, Q.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the State Key Laboratory of Geo-Information Engineering, grant numbers SKLGIE2022-ZZ2-03 and SKLGIE2023-M-2-1, and by the National Natural Science Foundation of China, grant number 42271164.

**Data Availability Statement:** Data can be found at https://github.com/duanxz0127/Co-viewContext (accessed on 27 October 2024).

**Acknowledgments:** The datasets for this experiment were made possible with the support of the WHU-TLS dataset and the generous assistance from Wuhan LuoJiaYiYun Optoelectronic Technology Co., Ltd. We express our heartfelt gratitude for their contributions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhang, J.; Singh, S. LOAM: Lidar Odometry and Mapping in Real-time. In Proceedings of the Robotics: Science and Systems X Robotics: Science and Systems Foundation, Berkeley, CA, USA, 12–16 July 2014. [CrossRef]
2. Shan, T.; Englot, B.; Meyers, D.; Wang, W.; Ratti, C.; Rus, D. LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 5135–5142. [CrossRef]
3. Xu, W.; Zhang, F. FAST-LIO: A Fast, Robust LiDAR-Inertial Odometry Package by Tightly-Coupled Iterated Kalman Filter. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3317–3324. [CrossRef]

4. Wang, Y.; Sun, Z.; Xu, C.-Z.; Sarma, S.E.; Yang, J.; Kong, H. LiDAR Iris for Loop-Closure Detection. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 5769–5775. [CrossRef]
5. Jiang, B.; Shen, S. Contour Context: Abstract Structural Distribution for 3D LiDAR Loop Detection and Metric Pose Estimation. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 8386–8392. [CrossRef]
6. Kim, G.; Kim, A. Scan Context: Egocentric Spatial Descriptor for Place Recognition Within 3D Point Cloud Map. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 4802–4809. [CrossRef]
7. Yuan, C.; Lin, J.; Zou, Z.; Hong, X.; Zhang, F. STD: Stable Triangle Descriptor for 3D place recognition 2023. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 1897–1903.
8. Kummerle, R.; Grisetti, G.; Strasdat, H.; Konolige, K.; Burgard, W. $G^2o$: A general framework for graph optimization. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3607–3613. [CrossRef]
9. Kaess, M.; Ranganathan, A.; Dellaert, F. iSAM: Incremental Smoothing and Mapping. *IEEE Trans. Robot.* **2008**, *24*, 1365–1378. [CrossRef]
10. Saeedi, S.; Trentini, M.; Seto, M.; Li, H. Multiple-Robot Simultaneous Localization and Mapping: A Review: Multiple-Robot Simultaneous Localization and Mapping. *J. Field Robot.* **2016**, *33*, 3–46. [CrossRef]
11. Zhou, X.; Roumeliotis, S. Multi-robot SLAM with Unknown Initial Correspondence: The Robot Rendezvous Case. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–13 October 2006; pp. 1785–1792. [CrossRef]
12. Gans, N.R.; Rogers, J.G. Cooperative Multirobot Systems for Military Applications. *Curr. Robot. Rep.* **2021**, *2*, 105–111. [CrossRef]
13. Wu, Z.; Pan, L.; Yu, M.; Liu, J.; Mei, D. A game-based approach for designing a collaborative evolution mechanism for unmanned swarms on community networks. *Sci. Rep.* **2022**, *12*, 18892. [CrossRef]
14. Queralta, J.P.; Taipalmaa, J.; Can Pullinen, B.; Sarker, V.K.; Nguyen Gia, T.; Tenhunen, H.; Gabbouj, M.; Raitoharju, J.; Westerlund, T. Collaborative Multi-Robot Search and Rescue: Planning, Coordination, Perception, and Active Vision. *IEEE Access* **2020**, *8*, 191617–191643. [CrossRef]
15. Drew, D.S. Multi-Agent Systems for Search and Rescue Applications. *Curr. Robot. Rep.* **2021**, *2*, 189–200. [CrossRef]
16. Alenzi, Z.; Alenzi, E.; Alqasir, M.; Alruwaili, M.; Alhmiedat, T.; Alia, O.M. A Semantic Classification Approach for Indoor Robot Navigation. *Electronics* **2022**, *11*, 2063. [CrossRef]
17. Alqobali, R.; Alshmrani, M.; Alnasser, R.; Rashidi, A.; Alhmiedat, T.; Alia, O.M. A Survey on Robot Semantic Navigation Systems for Indoor Environments. *Appl. Sci.* **2023**, *14*, 89. [CrossRef]
18. Leung, K.Y.K.; Barfoot, T.D.; Liu, H. Decentralized Localization of Sparsely-Communicating Robot Networks: A Centralized-Equivalent Approach. *IEEE Trans. Robot.* **2010**, *26*, 62–77. [CrossRef]
19. Besl, P.J.; McKay, N.D. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [CrossRef]
20. Chetverikov, D.; Stepanov, D.; Krsek, P. Robust Euclidean alignment of 3D point sets: The trimmed iterative closest point algorithm. *Image Vis. Vision. Comput.* **2005**, *23*, 299–309. [CrossRef]
21. Segal, A.V.; Hähnel, D.; Thrun, S. Generalized-ICP. In Proceedings of the Robotics: Science and Systems, Seattle, WA, USA, 28 June–1 July 2009; pp. 21–31.
22. Li, J.; Hu, Q.; Zhang, Y.; Ai, M. Robust symmetric iterative closest point. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 219–231. [CrossRef]
23. Hardouin, G.; Moras, J.; Morbidi, F.; Marzat, J.; Mouaddib, E.M. A Multirobot System for 3-D Surface Reconstruction with Centralized and Distributed Architectures. *IEEE Trans. Robot.* **2023**, *39*, 2623–2638. [CrossRef]
24. Takimoto, R.Y.; Tsuzuki, M.D.S.G.; Vogelaar, R.; Martins, T.D.C.; Sato, A.K.; Iwao, Y.; Gotoh, T.; Kagei, S. 3D reconstruction and multiple point cloud registration using a low precision RGB-D sensor. *Mechatronics* **2016**, *35*, 11–22. [CrossRef]
25. Liu, R.; Wang, J.; Zhang, B. High Definition Map for Automated Driving: Overview and Analysis. *J. Navig.* **2020**, *73*, 324–341. [CrossRef]
26. Li, Q.; Zhu, J.; Liu, J.; Cao, R.; Fu, H.; Garibaldi, J.M.; Li, Q.; Liu, B.; Qiu, G. 3D map-guided single indoor image localization refinement. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 13–26. [CrossRef]
27. Wu, H.; Yan, L.; Xie, H.; Wei, P.; Dai, J. A hierarchical multiview registration framework of TLS point clouds based on loop constraint. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 65–76. [CrossRef]
28. Wang, W.; Wang, B.; Zhao, P.; Chen, C.; Clark, R.; Yang, B.; Markham, A.; Trigoni, N. PointLoc: Deep Pose Regressor for LiDAR Point Cloud Localization. *IEEE Sens. J.* **2022**, *22*, 959–968. [CrossRef]
29. Charles, R.Q.; Su, H.; Kaichun, M.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85. [CrossRef]
30. Luo, L.; Zheng, S.; Li, Y.; Fan, Y.; Yu, B.; Cao, S.; Shen, H. BEVPlace: Learning LiDAR-based Place Recognition using Bird's Eye View Images 2023. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 8700–8709.

31. Wang, H.; Liu, Y.; Dong, Z.; Guo, Y.; Liu, Y.-S.; Wang, W.; Yang, B. Robust Multiview Point Cloud Registration with Reliable Pose Graph Initialization and History Reweighting. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 9506–9515. [CrossRef]
32. Yang, B.; Zang, Y.; Dong, Z.; Huang, R. An automated method to register airborne and terrestrial laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* **2015**, *109*, 62–76. [CrossRef]
33. Avidar, D.; Malah, D.; Barzohar, M. Local-to-Global Point Cloud Registration Using a Dictionary of Viewpoint Descriptors. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 891–899. [CrossRef]
34. Liang, F.; Yang, B.; Dong, Z.; Huang, R.; Zang, Y.; Pan, Y. A novel skyline context descriptor for rapid localization of terrestrial laser scans to airborne laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* **2020**, *165*, 120–132. [CrossRef]
35. Xu, D.; Liu, J.; Hyyppä, J.; Liang, Y.; Tao, W. A heterogeneous 3D map-based place recognition solution using virtual LiDAR and a polar grid height coding image descriptor. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 1–18. [CrossRef]
36. Ebadi, K.; Chang, Y.; Palieri, M.; Stephens, A.; Hatteland, A.; Heiden, E.; Thakur, A.; Funabiki, N.; Morrell, B.; Wood, S.; et al. LAMP: Large-Scale Autonomous Mapping and Positioning for Exploration of Perceptually-Degraded Subterranean Environments 2020. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020. pp. 80–86.
37. Chang, Y.; Ebadi, K.; Denniston, C.E.; Ginting, M.F.; Rosinol, A.; Reinke, A.; Palieri, M.; Shi, J.; Chatterjee, A.; Morrell, B.; et al. LAMP 2.0: A Robust Multi-Robot SLAM System for Operation in Challenging Large-Scale Underground Environments. *IEEE Robot. Autom. Lett.* **2022**, *7*, 9175–9182. [CrossRef]
38. Zhong, S.; Chen, H.; Qi, Y.; Feng, D.; Chen, Z.; Wu, J.; Wen, W.; Liu, M. CoLRIO: LiDAR-Ranging-Inertial Centralized State Estimation for Robotic Swarms 2024. *arXiv* **2024**, arXiv:2402.11790.
39. Kulkarni, M.; Dharmadhikari, M.; Tranzatto, M.; Zimmermann, S.; Reijgwart, V.; De Petris, P.; Nguyen, H.; Khedekar, N.; Papachristos, C.; Ott, L.; et al. Autonomous Teamed Exploration of Subterranean Environments using Legged and Aerial Robots. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 3306–3313. [CrossRef]
40. He, J.; Zhou, Y.; Huang, L.; Kong, Y.; Cheng, H. Ground and Aerial Collaborative Mapping in Urban Environments. *IEEE Robot. Autom. Lett.* **2021**, *6*, 95–102. [CrossRef]
41. Lajoie, P.-Y.; Ramtoula, B.; Chang, Y.; Carlone, L.; Beltrame, G. DOOR-SLAM: Distributed, Online, and Outlier Resilient SLAM for Robotic Teams. *IEEE Robot. Autom. Lett.* **2020**, *5*, 1656–1663. [CrossRef]
42. Arandjelovic, R.; Gronat, P.; Torii, A.; Pajdla, T.; Sivic, J. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1437–1451. [CrossRef]
43. Choudhary, S.; Carlone, L.; Nieto, C.; Rogers, J.; Christensen, H.I.; Dellaert, F. Distributed trajectory estimation with privacy and communication constraints: A two-stage distributed Gauss-Seidel approach. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 5261–5268. [CrossRef]
44. Mangelson, J.G.; Dominic, D.; Eustice, R.M.; Vasudevan, R. Pairwise Consistent Measurement Set Maximization for Robust Multi-Robot Map Merging. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Queensland, 21–25 May 2018; pp. 2916–2923. [CrossRef]
45. Huang, Y.; Shan, T.; Chen, F.; Englot, B. DiSCo-SLAM: Distributed Scan Context-Enabled Multi-Robot LiDAR SLAM With Two-Stage Global-Local Graph Optimization. *IEEE Robot. Autom. Lett.* **2022**, *7*, 1150–1157. [CrossRef]
46. Zhong, S.; Qi, Y.; Chen, Z.; Wu, J.; Chen, H.; Liu, M. DCL-SLAM: A Distributed Collaborative LiDAR SLAM Framework for a Robotic Swarm. *IEEE Sens. J.* **2024**, *24*, 4786–4797. [CrossRef]
47. Xu, Y.; Li, L.; Sun, S.; Wu, W.; Jin, A.; Yan, Z.; Yang, B.; Chen, C. Collaborative Exploration and Mapping with Multimodal LiDAR Sensors. In Proceedings of the 2023 IEEE International Conference on Unmanned Systems (ICUS), Hefei, China, 28–30 October 2023; pp. 1092–1097. [CrossRef]
48. Rusu, R.B. Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments. *Künstl Intell.* **2010**, *24*, 345–348. [CrossRef]
49. Zhang, W.; Qi, J.; Wan, P.; Wang, H.; Xie, D.; Wang, X.; Yan, G. An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. *Remote Sens.* **2016**, *8*, 501. [CrossRef]
50. Leutenegger, S.; Furgale, P.; Rabaud, V.; Chli, M.; Konolige, K.; Siegwart, R. Keyframe-Based Visual-Inertial SLAM using Nonlinear Optimization. In Proceedings of the Robotics: Science and Systems IX, Robotics: Science and Systems Foundation, Berlin, Germany, 24–28 June 2013. [CrossRef]
51. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [CrossRef]
52. LiDAR. Available online: http://luojiayy.com/product_detail_en/id/4.html (accessed on 27 March 2024).
53. Support for Matrice 300 RTK. Available online: https://www.dji.com/support/product/photo (accessed on 21 October 2024).
54. Alpha Uni 20: High-End LiDAR Solution, CHCNAV. Available online: https://chcnav.com/product-detail/alphauni-20 (accessed on 27 March 2024).
55. RIEGL—Produktdetail. Available online: http://www.riegl.com/nc/products/terrestrial-scanning/produktdetail/product/scanner/48/ (accessed on 27 March 2024).

56. Dong, Z.; Liang, F.; Yang, B.; Xu, Y.; Zang, Y.; Li, J.; Wang, Y.; Dai, W.; Fan, H.; Hyyppä, J.; et al. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *163*, 327–342. [CrossRef]
57. Livox Mid-360. Available online: https://www.livoxtech.com/cn/mid-360 (accessed on 27 October 2024).
58. Xu, W.; Cai, Y.; He, D.; Lin, J.; Zhang, F. FAST-LIO2: Fast Direct LiDAR-Inertial Odometry. *IEEE Trans. Robot.* **2022**, *38*, 2053–2073. [CrossRef]
59. Duan, X.; Hu, Q.; Zhao, P.; Yu, F.; Ai, M. A low-drift and real-time localisation and mapping method for handheld hemispherical view LiDAR-IMU integration system. *Photogramm. Rec.* **2023**, *38*, 176–196. [CrossRef]