전역 및 지역 동시성을 강건하게 보장하는 글로컬 융합 항법 시스템
Glocal Fusion for Navigation with Global and Local Consistency

초록

무인 이동체의 자율 임무 수행을 위해서는 다양한 센서를 융합하여 전역 및 지역 일관성이 동시에 보장되는 항법 시스템이 필요하다. 본 연구는 지역 SLAM이 제공하는 정밀한 상대 위치 추정과 누적 오차가 없는 전역 관측의 상호 보완적 특성을 활용하여 새로운 글로컬 융합 기법을 제안한다. 제안된 방법은 지역적으로 일관된 구간을 글로컬 그룹으로 묶고, 전역 관측은 개별 자세가 아닌 글로컬 상태를 보정하도록 구성하여 효과적으로 지역 일관성을 유지하면서 전역 일관성을 회복한다. 이를 기반으로 글로컬 관측기와 최적화기를 포함한 항법 시스템을 개발하였으며, 다양한 SLAM 및 전역 관측 실험을 통해 적은 변수로도 기존 방식보다 높은 수준의 일관성을 보임을 검증하였다.

Key Words : Robot navigation, sensor fusion, state estimation, factor graph optimization

1. Introduction

Robot navigation, which simultaneously maps the surrounding environment and estimates the robot's pose, plays an essential role in autonomous driving and robotic systems. To ensure safe and reliable operation, a navigation system should achieve both global and local consistency. Local consistency refers to the accuracy of relative pose estimation between the robot and its surrounding environment, whereas global consistency refers to the accuracy of pose estimation with respect to a global coordinate frame.

Over the past three decades, local simultaneous localization and mapping (SLAM) methods have been extensively studied using vision[1~3] or LiDAR sensors[4~6], which provide environmental observations. Local SLAM systems can achieve highly precise pose estimation and environmental mapping within local regions. However, due to accumulated estimation errors, local SLAM inherently suffers from global inconsistency over time. Although loop closure modules have been proposed to address this issue, they require the robot to revisit previously explored areas, which imposes operational constraints and demands continuous resources for loop detection.

To address the limitations of local SLAM and ensure both global and local consistency during long-term operation, research has focused on sensor fusion frameworks that incorporate global observations. Unlike local SLAM, global observations are free from drift. However, global observations obtained from low-cost sensors commonly used in robotic navigation are often noisy and prone to unexpected outliers, making them unreliable for direct use in navigation. Therefore, it is essential to exploit the complementary characteristics of precise but drift-prone local observations and noisy but drift-free global observations, facilitating the development of sensor fusion systems that ensure both local and global consistency.

Numerous studies have explored sensor fusion using various sensor modalities. Approaches for state estimation can be categorized into filtering-based methods[7~9] and optimization-based methods[10~13]. Recently, the latter has been increasingly adopted due to its higher estimation accuracy enabled by iterative relinearization. These methods typically optimize either individual local poses or local poses together with global states that represent the transformation between the local and global world frames. As a result, a factor graph is typically constructed as shown in Fig. 1, and state estimation is performed by optimizing it.

However, optimizing such factor graphs poses several challenges. First, these graphs have a structure in which global observations directly affect individual local poses, which may degrade the well-maintained local consistency due to the high noise or outliers present in global data. Second, they are also characterized by heterogeneous factors from global and local observations constraining the same state simultaneously. In this case, the estimation performance becomes highly sensitive to the accuracy of the assigned covariance models. Yet, in local SLAM, computing reliable pose covariance in real time is often impractical due to computational constraints, so fixed

covariance models are commonly used instead[10]. The discrepancy between the actual and assigned covariances can result in degraded sensor fusion performance.

We propose a novel glocal fusion approach to address these challenges. It is analyzed that drift in local SLAM does not accumulate steadily but remains well constrained until certain regions where local maps become corrupted, as shown in Fig. 2. To exploit this property, a new parameterization is introduced that groups locally consistent segments into glocal groups, each associated with a glocal state for correcting global consistency. The factor graph is then structured so that global factors influence only the glocal states, without directly affecting the locally consistent local states within each group, as shown in Fig. 3. This design allows noisy global data to recover global consistency without degrading local consistency, and mitigates the risks inherent in conventional factor graphs by decoupling tightly connected factors.

A glocal navigation system is proposed to implement the proposed glocal fusion approach in a practical online system. This system consists of two main modules. The first is the glocal observer, which evaluates the local consistency of poses using global observations and groups locally consistent segments into glocal groups. The second is the glocal optimizer, which corrects global consistency of the identified glocal groups and ensures smooth transitions between them.

The contributions of this paper are summarized as follows:

- A novel glocal fusion approach is proposed to effectively exploit the complementary characteristics of local and global data. The proposed approach enables both improved optimization efficiency, through reduced state variables, and better preservation of global and local consistency compared to conventional fusion approaches. This advantage is systematically discussed at the end of Section 4.3.
- The proposed glocal fusion approach is formally established, from the introduction of a novel parameterization to its formulation based on maximum a posteriori (MAP) estimation, providing a general framework applicable to various combinations of global and local sensor modalities.
- The glocal navigation system is developed based on the proposed glocal fusion approach. The system comprises a glocal observer module, which segments locally consistent regions, and a glocal optimizer module, which jointly optimizes glocal states and local states to maintain global consistency and smooth transitions between groups.
- Extensive experiments are conducted using various local SLAM algorithms and global observations. The results show that the proposed approach achieves superior preservation of both global and local consistency with significantly fewer parameters compared to conventional fusion systems.

The remainder of this paper is structured as follows. Section 2 reviews existing sensor fusion approaches, and Section 3 introduces the mathematical background. The proposed glocal fusion approach is presented in Section 4, including the analysis of local SLAM characteristics (Section 4.1), the introduction of a new parameterization (Section 4.2), and its formulation based on the MAP estimation (Section 4.3). The overview of the glocal navigation system is provided in Section 5.1, followed by detailed descriptions of the glocal observer (Section 5.2) and glocal optimizer modules (Section 5.3). Finally, experimental results and conclusions are presented in Sections 6 and 7, respectively.

2. Related Work

This section reviews related works on sensor fusion with various global and local observations, as well as approaches that manage factor graphs in a submap-wise manner.

2.1. Sensor Fusion with Global and Local Data

In robotics, common sources of global information include global navigation satellite systems (GNSS), indoor positioning systems based on Wi-Fi or ultra-wideband (UWB), and localization with respect to a prior map.

In outdoor environments, GNSS is the most widely used source, providing global positions (and velocities) in the Earth-Centered Earth-Fixed (ECEF) coordinate frame. Filtering-based methods[7-9] have fused GNSS with local sensors such as vision, IMU, or LiDAR in tightly or loosely coupled manners to estimate individual local poses. Factor

graph optimization (FGO)-based approaches[10-11] have also been used in loosely coupled manners, integrating GNSS position solutions with odometry estimates from local sensors. VINS-Fusion[10], in particular, presents a general framework applicable to various sensor modalities including GNSS. Tightly coupled fusion of precise GNSS RTK solutions with visual features and IMU data has been proposed[12]. In the other work[13], instead of using GNSS solutions, raw observations including pseudorange and Doppler shift are directly integrated through FGO to estimate both local poses and a global state, which represents the transformation between the local world frame of the visual inertial odometry (VIO) and the ECEF frame.

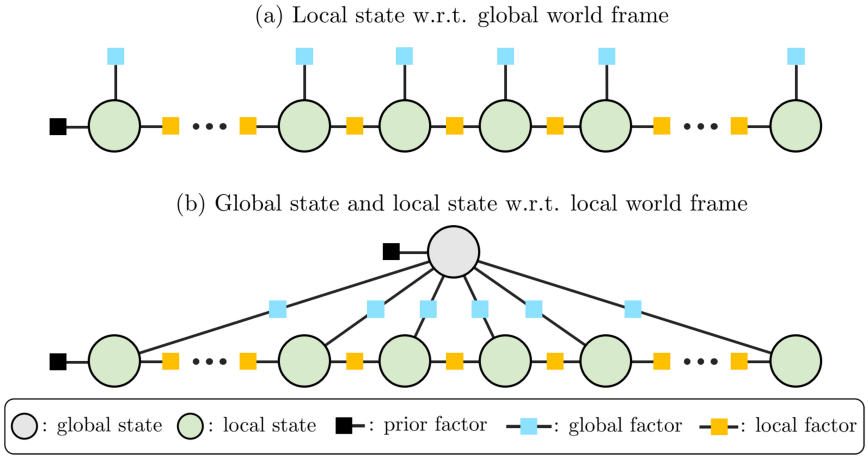

Fig. 1. Factor graph representations of conventional fusion approaches.

Indoor positioning can be performed using signals from anchor stations, such as WiFi or UWB, providing positions relative to a predefined global frame. UWB-based positioning[14] and WiFi-based positioning[16] have both been integrated with local estimation using FGO. Furthermore, tightly coupled fusion of visual-inertial and UWB observations using FGO has been investigated[17,18] for joint estimation of poses and anchor station locations.

When a prior map is available, localization within the global frame defined by the map enables the estimation of global poses. Various map representations have been used, such as dense 3D point clouds[19-22], surfel maps[23], and the map based on Gaussian mixture models[24]. Localization is typically performed by registering a local map, built by a SLAM or odometry system, to the prior map. FGO-based approaches[19,20] fuse local odometry with global localization to reduce pose drift, while the work[24] incorporate global registration residuals into local bundle adjustment, improving both local map accuracy and pose estimates.

Conventional sensor fusion approaches commonly employ factor graphs where individual local poses are treated as state variables, as illustrated in Fig. 1. However, this approach may suffer from the structural risks discussed in Section 1. The proposed glocal fusion approach introduces a new parameterization that leverages the complementary characteristics of local and global sensors to overcome these limitations. To the best of our knowledge, this is the first study that evaluates local SLAM consistency with global observations, groups locally consistent regions, and performs efficient optimization based on such grouping within a sensor fusion framework.

2.2. Submap Partitioning

Partitioning a large factor graph into submaps and performing hierarchical optimization[25,26] is a widely adopted strategy for scalability, especially in long-term navigation in large environments[27-34]. Typically, this hierarchical framework optimizes states within each submap at the lower level and aligns the submaps at the higher level to maintain consistency while reducing computational complexity.

The method of partitioning submaps plays a critical role in submap-based strategies. To maintain balanced submap sizes and information content, works[28-30] partition submaps when the accumulated travel distance, number of mapped points, or number of scans exceeds a predefined threshold. To ensure local consistency within submaps, studies[31,32] employ partitioning criteria based on pose covariance estimates or local factor residuals. To preserve independence between submaps, works[33,34] partition the graph at points where covisible data shared between submaps is minimized. Graph-theoretic approaches, such as spectral graph partitioning and normalized cuts[26,27],

are also applied to identify boundaries that maximize intra-submap connectivity while minimizing inter-submap connections.

While existing methods rely on predefined parameters or community detection within factor graphs, they typically depend solely on local information, without incorporating global observations. Consequently, they may fail to detect segments with degraded local consistency or generate excessive redundant submaps. In contrast, the proposed glocal fusion approach leverages global observations via a dedicated glocal observer module, which enables accurate segmentation of locally consistent regions and efficient group-wise optimization.

3. Preliminaries

The proposed glocal fusion system utilizes states defined on Riemannian manifolds, specifically the Special Orthogonal Group $\mathrm{SO}(3)$, the Special Euclidean Group $\mathrm{SE}(3)$, and the Similarity Transformation Group $\mathrm{SIM}(3)$. Prior to presenting the proposed methodology, this section outlines key concepts of Lie groups and Lie algebras, based on the paper[35].

3.1. Exponential and Logarithmic Maps

A Lie group $\mathcal{G}$ is a smooth manifold with a group structure. Its associated Lie algebra $\mathfrak{g}$, defined as the tangent space at the identity element, provides a linearized representation of the local structure of $\mathcal{G}$. The group and its algebra are related via the exponential map $\exp\colon \mathfrak{g} \to \mathcal{G}$ and the logarithmic map $\log\colon \mathcal{G} \to \mathfrak{g}$, both locally defined around the identity. For the Lie groups considered in this work, $\mathfrak{g}$ is isomorphic to a vector space $\mathbb{R}^n$. The mapping between $\mathbb{R}^n$ and $\mathfrak{g}$ is defined by $(\cdot)^\wedge : \mathbb{R}^n \to \mathfrak{g}$ and $(\cdot)^\vee : \mathfrak{g} \to \mathbb{R}^n$. For notational convenience, the mappings between $\mathbb{R}^n$ and $\mathcal{G}$ are denoted as follows: $\mathrm{Exp}\colon \mathbb{R}^n \to \mathcal{G}$ and $\mathrm{Log}\colon \mathcal{G} \to \mathbb{R}^n$.

3.2. Adjoint Map, Jacobian, and Uncertainty

Since the Lie groups considered in this paper are non-commutative, the adjoint map is employed to account for the effect of group multiplication when representing Lie algebra elements in vector form. The adjoint map describes how a perturbation in the Lie algebra is transformed under conjugation by the group element. The vector-space adjoint map is defined as: $\mathrm{Ad}(\mathbf{G})\mathbf{v} := (\mathbf{G}\hat{\mathbf{v}}\mathbf{G}^{-1})^\vee$, where $\mathbf{G} \in \mathcal{G}$ and $\mathbf{v} \in \mathbb{R}^n$. Using the definition, the following identities can be derived:

$$\begin{aligned} \mathbf{G} \cdot \mathrm{Exp}(\mathbf{v}) &= \mathrm{Exp}(\mathrm{Ad}(\mathbf{G})\mathbf{v}) \cdot \mathbf{G}, \\ \mathrm{Exp}(\mathbf{v}) \cdot \mathbf{G} &= \mathbf{G} \cdot \mathrm{Exp}(\mathrm{Ad}(\mathbf{G})^{-1}\mathbf{v}). \end{aligned} \tag{1}$$

Since Lie groups are nonlinear manifolds, linearization in the associated Lie algebras is essential for optimization and uncertainty propagation. The composition of perturbations on Lie groups can be locally approximated by the first-order Baker-Campbell-Hausdorff (BCH) formula[36]:

$$\mathrm{Exp}(\mathbf{v}_1)\,\mathrm{Exp}(\mathbf{v}_2) \approx \begin{cases} \mathrm{Exp}\left(\mathbf{J}_1^{-1}(\mathbf{v}_2)\,\mathbf{v}_1 + \mathbf{v}_2\right), \text{small } \mathbf{v}_1 \\ \mathrm{Exp}\left(\mathbf{v}_1 + \mathbf{J}_\mathrm{r}^{-1}(\mathbf{v}_1)\,\mathbf{v}_2\right), \text{small } \mathbf{v}_2 \end{cases}. \tag{2}$$

Uncertainty on Lie groups can be approximated using a small perturbation defined in the tangent space at the identity. Throughout this paper, we adopt the right perturbation convention: $\mathbf{G}^* := \mathbf{G} \cdot \mathrm{Exp}(\delta\mathbf{v})$, $\delta\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Here, $\mathbf{G}^*$ denotes the true state, and the uncertainty of the estimate $\mathbf{G}$ is represented by a right perturbation $\delta\mathbf{v}$.

3.3. Notations of Lie Groups

The group $\mathrm{SO}(3)$ represents 3D rotations, while $\mathrm{SE}(3)$ describes 3D rigid-body transformations. The group $\mathrm{SIM}(3)$ further extends $\mathrm{SE}(3)$ by incorporating a scale factor, thereby representing similarity transformations. The notations

for the groups and their corresponding vectors are defined as follows:

$$\begin{aligned} &\mathtt{R} \in \mathrm{SO}(3), && \boldsymbol{\omega} = \mathrm{Log}(\mathtt{R}) \in \mathbb{R}^3, \\ &\mathtt{T} = \begin{bmatrix} \mathtt{R} & \mathsf{p} \\ \mathsf{0} & 1 \end{bmatrix} \in \mathrm{SE}(3), && \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\omega} \\ \boldsymbol{\nu} \end{bmatrix} = \mathrm{Log}(\mathtt{T}) \in \mathbb{R}^6, \\ &\mathtt{S} = \begin{bmatrix} s\mathtt{R} & \mathsf{p} \\ \mathsf{0} & 1 \end{bmatrix} \in \mathrm{SIM}(3), && \boldsymbol{\eta} = \begin{bmatrix} \lambda \\ \boldsymbol{\omega} \\ \boldsymbol{\nu} \end{bmatrix} = \mathrm{Log}(\mathtt{S}) \in \mathbb{R}^7, \end{aligned} \tag{3}$$

where $s = e^{\lambda} \in \mathbb{R}$, $\mathsf{p} \in \mathbb{R}^3$, $\boldsymbol{\nu} \in \mathbb{R}^3$. Here, λ, $\boldsymbol{\omega}$, and $\boldsymbol{\nu}$ denote the scale, rotation, and translation components of the vector representation, respectively. Detailed formulations of the exponential and logarithmic maps, adjoint map, and Jacobians for each Lie group can be found in the paper[37].

4. Glocal Fusion Approach

In this section, the characteristics of the local SLAM system are first analyzed, and based on this analysis, the concepts of glocal groups and glocal states are defined to introduce a novel parameterization. Additionally, the proposed glocal state estimation is derived from a MAP perspective.

4.1. Local Navigation Characteristics

This subsection analyzes a local SLAM system to identify key characteristics for the novel parameterization of the glocal fusion approach.

A typical SLAM system comprises two main components: a frontend that constructs a factor graph by associating current observations with the local map, and a backend that optimizes this graph using local factors, which impose constraints between pose and map nodes. A local factor constraining the pose at time k and the j-th map point is formulated as follows:

$$\begin{aligned} \tilde{\boldsymbol{\ell}}_{kj} &= \mathsf{f}_{\ell}\left(\mathrm{Warp}\left(\mathtt{T}_{wk}, \mathsf{m}_{wj}\right)\right) + \mathsf{n}_{kj}, \\ \text{where } \ \mathrm{Warp}\left(\mathtt{T}_{wk}, \mathsf{m}_{wj}\right) &= \mathtt{R}_{wk}^{\top}\left(\mathsf{m}_{wj} - \mathsf{p}_{wk}\right) = \mathsf{m}_{kj}. \end{aligned} \tag{4}$$

Here, $\tilde{\boldsymbol{\ell}}_{kj}$ denotes the observation of the j-th map point obtained from the pose at time k, and n_{kj} represents the observation noise. The notations $\mathtt{T}_{wk}$, $\mathtt{R}_{wk}$, and p_{wk} represent the 3D pose, rotation, and translation of the sensor frame at time k with respect to the world frame $\{w\}$, respectively. The point m_{wj} is the j-th map point in the world frame, and $\mathsf{f}_{\ell}(\mathsf{m})$ maps a 3D point into the observation space; for example, it becomes a projection function to image coordinates $[u, v]^{\top}$ in monocular SLAM, or the identity function in LiDAR SLAM.

The factor graph of a local SLAM consists of pose and map point nodes connected by numerous local factors, and the FGO is used to estimate both poses and map points. However, during optimization, there exists an unobservable subspace that cannot be constrained by the local factors alone, referred to as gauge freedomfigure38. This concept can be intuitively understood using a LiDAR-based SLAM system. When all pose and map point nodes in the world frame $\{w\}$ are transformed to an arbitrary frame $\{a\}$ using $\mathtt{T}_{aw}$, their relative geometry remains unchanged. As a result, the local factors are invariant, and the system has six degrees of gauge freedom, corresponding to $\mathrm{SE}(3)$. In monocular SLAM, the projection function $\mathsf{f}_{\ell}(\cdot)$ in (4) makes the absolute scale unobservable, introducing an additional degree of freedom. Therefore, the system exhibits seven degrees of freedom, corresponding to $\mathrm{SIM}(3)$figure39.

In real-time SLAM systems, sliding-window optimization is commonly employed, where only the nodes within a fixed-size window are retained in the local factor graph and optimized. When optimization relies solely on local factors, gauge freedom drift may occur, causing inconsistency between current nodes and old nodes outside the window. To suppress such drift, old nodes are marginalized out and replaced with prior factors that encode their constraints with the retained nodes. In typical settings, dense local connectivity ensures that these prior factors are sufficiently strong to prevent drift. However, unexpected scenarios such as sudden changes in environmental observations or failures in the data association can still lead to gauge freedom drift.

For instance, during a rotation followed by straight-line motion, observations obtained after the rotation may exhibit strong mutual overlap but limited overlap with pre-rotation observations. In such cases, the information from local factors within the current window can significantly outweigh that of the prior, causing the current nodes to

converge to a solution that maintains local consistency but deviates from global consistency. As a result, the current local map becomes corrupted along gauge freedom directions relative to the old nodes. Subsequently, strong local consistency is re-established within this distorted local map.

The characteristics of the local SLAM can be summarized as follows. Relative pose errors do not accumulate gradually over time but remain very small throughout most of the trajectory due to strong local consistency. In certain segments with weak data association, however, global consistency can be abruptly corrupted along directions of the system's gauge freedom.

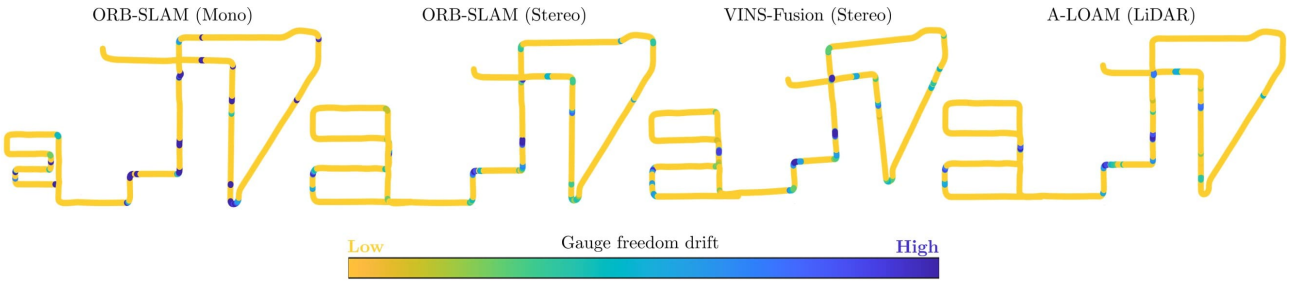

Fig. 2. Visualization of gauge freedom drift across four types of local SLAM systems: monocular and stereo versions of ORB-SLAM[1], the stereo version of VINS-Fusion[10], and A-LOAM[4], tested on sequence 08 of the KITTI dataset[40]. The parameters used for computation are a traveling length of $10\,\mathrm{m}$ and a confidence level $\beta = 1-\sigma$ (approximately 68%). The standard deviations for local SLAM are $0.05 \times \frac{\pi}{180}$ radians for rotation and $0.05\,\mathrm{m}$ for translation.

4.2. Glocal Group and State Representation

This subsection introduces a novel state parameterization for the proposed fusion system, designed to reflect the local SLAM characteristics identified in Section 4.1.

When local consistency is well maintained within a specific segment, the estimated relative poses remain accurate, regardless of the segment's global consistency. In such cases, correcting the global misalignment of the segment automatically restores the global consistency of all included poses, without the need to adjust each pose individually. Based on this observation, we propose to group temporally adjacent local poses that exhibit strong local consistency into a single unit, referred to as a glocal group:

$$\begin{aligned} \mathcal{K}_i := \{k \mid & \delta\xi^{\ell}_{ek} = \mathrm{Log}\left(\delta\mathrm{T}^{\ell}_{ek}\right) \sim \mathcal{N}\left(\mathbf{0}, \Sigma^{\ell}\right), \\ & \delta\mathrm{T}^{\ell}_{ek} = \mathcal{F}\left(\mathrm{X}_{ew_i}, \tilde{\mathrm{T}}^{\ell}_{w_i k}\right)^{-1} \mathrm{T}^{*}_{ek}, \\ & (k-1 \in \mathcal{K}_i \text{ or } k+1 \in \mathcal{K}_i)\}. \end{aligned} \tag{5}$$

Here, $\mathcal{K}_i$ denotes the index set of local poses belonging to the i-th glocal group. The frames $\{e\}$ and $\{w_i\}$ represent the global world frame and the i-th glocal world frame, respectively. The pose $\tilde{\mathrm{T}}^{\ell}_{w_i k}$ indicates the local SLAM estimate at time k, expressed in $\{w_i\}$, and is obtained by transforming the SLAM output $\tilde{\mathrm{T}}^{\ell}_{wk}$ in $\{w\}$ via $\tilde{\mathrm{T}}^{\ell}_{w_i k} = \mathrm{T}_{w_i w}\,\tilde{\mathrm{T}}^{\ell}_{wk}$. The frame $\{w_i\}$ is typically initialized as the pose of the first local pose in the group. A correction function $\mathcal{F}(\cdot)$ is applied to the group using the group-level parameter X_{ew_i} to mitigate global drift. A pose is considered locally consistent and eligible for inclusion in the group if the global error $\delta\mathrm{T}^{\ell}_{ek}$ between the corrected pose and the ground-truth pose T^{*}_{ek} follows the distribution defined by the local SLAM covariance Σ^{ℓ}.

Intuitively, the glocal group definition in (5) implies that a single group-level correction parameter X_{ew_i} is sufficient to reduce the global errors of local poses within the group. The local SLAM covariance Σ^{ℓ} serves as a tunable threshold that defines the acceptable bounds for local consistency: a larger Σ^{ℓ} allows more poses to be grouped with weaker consistency, while a smaller value enforces stricter consistency, resulting in smaller glocal groups.

The correction function $\mathcal{F}(\cdot)$ and the group-level correction parameter $\mathtt{X}_{ew_i}$ in (5) are defined as follows:

$$\mathcal{F}\left(\mathtt{X}_{ew_i}, \tilde{\mathtt{T}}^{\ell}_{w_i k}\right) := \begin{cases} \mathtt{X}_{ew_i}\ \tilde{\mathtt{T}}^{\ell}_{w_i k}, & \mathtt{X}_{ew_i} \in \mathrm{SE}(3) \\ \mathtt{X}_{ew_i}\ \tilde{\mathtt{T}}^{\ell}_{w_i k}\ \underline{\mathtt{X}_{ew_i}}, & \mathtt{X}_{ew_i} \in \mathrm{SIM}(3) \end{cases}, \tag{6}$$
$$\text{where } \underline{\mathtt{X}_{ew_i}} = \begin{bmatrix} \frac{1}{s_{ew_i}}\mathtt{I} & \mathtt{0} \\ \mathtt{0} & 1 \end{bmatrix}.$$

The group-level correction parameter $\mathtt{X}_{ew_i}$ is referred to as the glocal state. The glocal state is defined on the gauge freedom of the local SLAM system: $\mathrm{SIM}(3)$ for monocular SLAM, and $\mathrm{SE}(3)$ for stereo or LiDAR-based SLAM. Since the corrected pose must lie on $\mathrm{SE}(3)$, the scale component of $\mathtt{X}_{ew_i} \in \mathrm{SIM}(3)$ is canceled via right multiplication with $\underline{\mathtt{X}_{ew_i}}$.

By defining the glocal state along the gauge freedom of the local SLAM system, the correction process for global consistency does not interfere with the strong local consistency maintained by local factors. This is because the glocal state lies entirely within the unobservable directions. Consequently, the proposed parameterization of glocal groups and glocal states enables global consistency to be restored while preserving the high local consistency inherent to the original SLAM system.

Fig. 2 illustrates the key characteristics of local SLAM analyzed in Section 4.1 using the proposed parameterization. The gauge freedom drift ϵ_k is visualized for four representative SLAM systems, where the drift at the k-th pose is defined as:

$$\epsilon_k = \begin{cases} \|\mathrm{Log}\left(\mathtt{X}_{eb}(k)^{-1}\,\mathtt{X}_{ef}(k)\right)\|_2, & E_k \geq \mathrm{CDF}^{-1}_{\chi^2_6}(\beta) \\ 0, & E_k < \mathrm{CDF}^{-1}_{\chi^2_6}(\beta) \end{cases}, \tag{7}$$
$$\text{where } \mathtt{X}_{e\square}(k) = \underset{\mathtt{X}_{ek}}{\mathrm{argmin}} \sum_{k' \in \mathcal{K}_\square} \|\boldsymbol{\gamma}_{kk'}\|^2_{\Sigma^\ell},$$
$$E_k = \frac{1}{\mathrm{Card}\left(\mathcal{K}_a\right)} \min_{\mathtt{X}_{ek}} \sum_{k' \in \mathcal{K}_a} \|\boldsymbol{\gamma}_{kk'}\|^2_{\Sigma^\ell},$$

$\boldsymbol{\gamma}_{kk'} = \mathrm{Log}\left(\mathcal{F}\left(\mathtt{X}_{ek}, \tilde{\mathtt{T}}^{\ell}_{kk'}\right)^{-1}\mathtt{T}^{*}_{ek'}\right)$, and $\|\boldsymbol{\gamma}\|^2_{\Sigma} = \boldsymbol{\gamma}^\top \Sigma^{-1}\boldsymbol{\gamma}$. Here, the symbol $\square$ indicates either f (forward) or b (backward), relative to the reference index k. The sets $\mathcal{K}_f$ and $\mathcal{K}_b$ represent forward and backward glocal groups, respectively, containing poses located after and before the k-th pose, such that the cumulative traveled distance remains below a predefined threshold. Their union defines $\mathcal{K}_a = \mathcal{K}_f \cup \mathcal{K}_b$. E_k quantifies the local consistency of $\mathcal{K}_a$, following a chi-squared distribution. If E_k falls within the confidence threshold β (i.e., $E_k < \mathrm{CDF}^{-1}_{\chi^2_6}(\beta)$), the group is deemed locally consistent and ϵ_k is set to zero. Otherwise, ϵ_k is computed as the relative glocal state $\mathtt{X}_{eb}(k)^{-1}\mathtt{X}_{ef}(k)$, representing the gauge freedom drift between the forward and backward groups.

As shown in Fig. 2 and analyzed in Section 4.1, all tested local SLAM systems maintain strong local consistency throughout most of the trajectory, resulting in negligible gauge freedom drift. However, in certain segments such as during sharp rotations, gauge freedom drift arises and corrupts global consistency. If such drift occur at a specific point, each local pose can be partitioned into glocal groups based on the definition in (5), and corrected using only the corresponding glocal state. However, gauge freedom drift does not occur instantaneously at a single pose. Instead, due to the nature of sliding-window optimization, the drift gradually propagates across multiple local poses. To correct this propagated drift between adjacent glocal groups, the proposed approach retains a subset of local poses near each group boundary as additional optimization variables. Accordingly, the set of optimization variables in the proposed glocal fusion framework consists of the glocal states and the boundary local poses that bridge adjacent groups, and is modeled as follows:

$$\mathtt{T}^{*}_{ek} = \begin{cases} \mathcal{F}\left(\mathtt{X}_{ew_i}\delta\mathtt{X}_{ew_i},\ \tilde{\mathtt{T}}^{\ell}_{w_i k}\right)\mathrm{Exp}\left(\delta\mathsf{n}^{\ell}_{k}\right), & (k \in \mathcal{K}^{\mathrm{h}}_i) \\ \mathcal{F}\left(\mathtt{X}_{ew_i}\delta\mathtt{X}_{ew_i},\ \mathtt{T}_{w_i k}\delta\mathtt{T}_{w_i k}\right), & (k \in \mathcal{K}^{\mathrm{t}}_i) \end{cases}, \tag{8}$$

where $\delta\mathsf{n}^{\ell}_{k} \sim \mathcal{N}\left(\mathtt{0}, \Sigma^{\ell}\right)$. Here, $\delta\mathtt{X}_{ew_i}$ and $\delta\mathtt{T}_{w_i k}$ denote the updates to the optimization variables corresponding to the glocal state and the local pose, respectively. The sets $\mathcal{K}^{\mathrm{h}}_i$ and $\mathcal{K}^{\mathrm{t}}_i$ are disjoint subsets of the glocal group $\mathcal{K}_i$, where $\mathcal{K}^{\mathrm{t}}_i$ contains the last N_ℓ poses, and $\mathcal{K}^{\mathrm{h}}_i$ includes the remaining poses. For two adjacent glocal groups, the tail subset of the preceding group plays a key role in correcting the gradual gauge freedom drift that propagates across the boundary. For the last glocal group, $\mathcal{K}^{\mathrm{t}}_i$ is defined as the empty set.

Compared to conventional fusion approaches that treat all local poses as optimization variables, the proposed parameterization exploits the observed characteristics of local SLAM to significantly reduce the number of variables while maintaining correction effectiveness. Furthermore, since the glocal state is defined on the unobservable gauge freedom subspace of the local factor graph, applying global corrections via the glocal state does not affect the well-maintained local consistency.

4.3. Maximum a Posteriori Glocal State Estimation

This subsection formulates the proposed glocal fusion approach from a MAP estimation perspective, based on the novel parameterization introduced in Section 4.2.

To aid understanding, a simplified factor graph is considered, as shown in Fig. 3(a), where the index set of local poses $\mathcal{K}$ is partitioned into two glocal groups: $\mathcal{K}_i = \{k_1^i, \cdots, k_n^i\}$ and $\mathcal{K}_j$. The group $\mathcal{K}_i$ is further divided into head and tail subsets: $\mathcal{K}_i^{\mathrm{h}} = \mathcal{K}_i \setminus \mathcal{K}_i^{\mathrm{t}}$ and $\mathcal{K}_i^{\mathrm{t}} = \{k_{n-N_\ell+1}^i, \cdots, k_n^i\}$. Since the global observations $\{\mathtt{G}_k\}$ are defined in the global frame, they are modeled using both the glocal state and the associated local poses. Local observations $\{\mathtt{L}_{k-1,k}\}$ are represented as binary factors between temporally adjacent poses. For clarity, the glocal state and local pose are denoted by $\mathtt{X}_i$ and $\mathtt{T}_k$, respectively. The derivation below focuses on the case $\mathtt{X}_i \in \mathrm{SE}(3)$, while the formulation naturally extends to $\mathrm{SIM}(3)$.

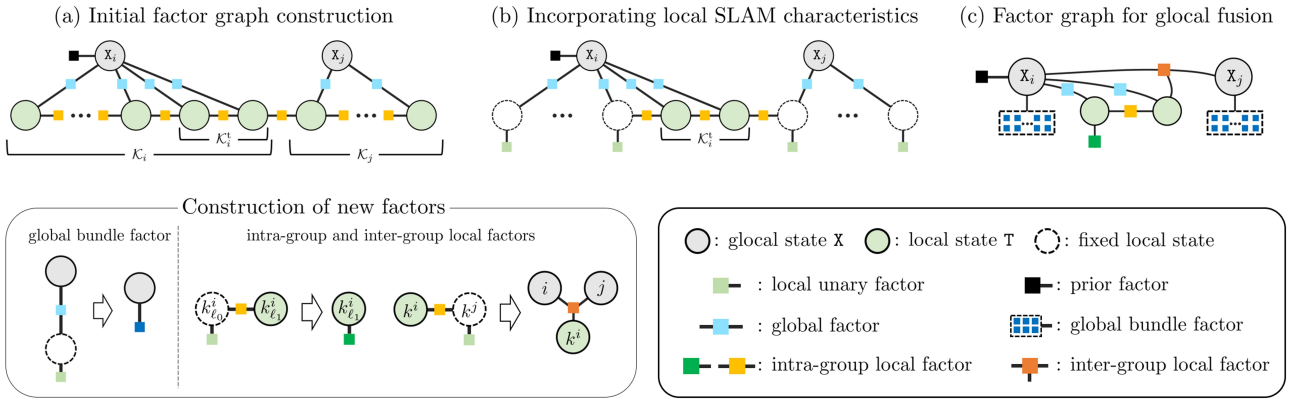

Fig. 3. Factor graph representation of the proposed glocal fusion approach. (a) Initial factor graph constructed with global and local factors. (b) Incorporation of local SLAM characteristics by treating strongly constrained local states as fixed random variables. (c) Final factor graph for the glocal fusion approach.

Let the observation and initial state sets be defined as $\mathcal{Z} := \{\{\mathtt{G}_k\}_{k\in\mathcal{K}}, \{\mathtt{L}_{k-1,k}\}_{(k-1,k)\in\mathcal{K}}\}$ and $\mathcal{X} := \{\mathtt{X}_i, \mathtt{X}_j, \{\mathtt{T}_k\}_{k\in\mathcal{K}}\}$, respectively. The posterior distribution of $\mathcal{X}$ can be factorized into prior and likelihood terms by applying Bayes' rule, the Markov property, and the conditional independence of observations given those states:

$$p(\mathcal{X} \mid \mathcal{Z}) \propto p(\mathtt{X}_i) \prod_{k\in\mathcal{K}} p(\mathtt{G}_k \mid \mathtt{X}_i, \mathtt{X}_j, \mathtt{T}_k) \prod_{(k-1,k)\in\mathcal{K}} p(\mathtt{L}_{k-1,k} \mid \mathtt{X}_i, \mathtt{X}_j, \mathtt{T}_{k-1}, \mathtt{T}_k). \tag{9}$$

The likelihood terms associated with local observations can be decomposed based on the glocal group as follows:

$$\prod_{k\in\mathcal{K}_i^{\mathrm{h}}} p(\mathtt{L}_k \mid \mathtt{T}_k) \prod_{k\in\mathcal{K}_i^{\mathrm{t}}\cup\{k_1^j\}} p(\mathtt{L}_{k-1,k} \mid \mathtt{T}_{k-1}, \mathtt{T}_k) \prod_{k\in\mathcal{K}_j} p(\mathtt{L}_k \mid \mathtt{T}_k). \tag{10}$$

Due to the definition of the glocal group, no local factors connecting $\mathcal{K}_i^{\mathrm{h}}$ and $\mathcal{K}_j$ constrain the glocal states $\mathtt{X}_i$ or $\mathtt{X}_j$. Therefore, the corresponding likelihood terms depend only on the local states $\mathtt{T}_k$. Similarly, the global observation likelihood terms in (9) can be partitioned by glocal group as:

$$\prod_{k\in\mathcal{K}_i^{\mathrm{h}}} p(\mathtt{G}_k \mid \mathtt{X}_i, \mathtt{T}_k) \prod_{k\in\mathcal{K}_i^{\mathrm{t}}} p(\mathtt{G}_k \mid \mathtt{X}_i, \mathtt{T}_k) \prod_{k\in\mathcal{K}_j} p(\mathtt{G}_k \mid \mathtt{X}_j, \mathtt{T}_k). \tag{11}$$

The local states $\{\mathtt{T}_k\}_{k\in\mathcal{K}_i^{\mathrm{h}}\cup\mathcal{K}_j}$ preserve strong local consistency due to the presence of numerous local factors. However, global factors often have higher uncertainty and may include outliers, potentially degrading this consis-

tency. To mitigate such risk, the local states in $\mathcal{K}_i^{\mathrm{h}} \cup \mathcal{K}_j$ are modeled as fixed random variables[1] to retain their consistency. Their uncertainty is modeled using the local SLAM covariance employed in the glocal group definition (5). Since local states within the same group satisfy this covariance condition by definition, they are treated as fixed random variables with the corresponding covariance. When a new glocal group is formed, the last N_ℓ poses from the preceding group are promoted to optimization variables to correct drift propagated across group boundaries. The corresponding factor graph is shown in Fig. 3(b).

Based on the fixed local states, the group-wise likelihood can be reformulated. First, the global observation likelihood associated with fixed local states ($k \in \mathcal{K}_i^{\mathrm{h}} \cup \mathcal{K}_j$) becomes a unary factor depending only on the glocal state. This term is referred to as the global bundle factor:

$$p\left(\mathtt{G}_k \mid \mathtt{X}_i, \mathtt{T}_k\right) p\left(\mathtt{L}_k \mid \mathtt{T}_k\right) \propto p\left(\mathtt{G}_k, \mathtt{L}_k \mid \mathtt{X}_i\right). \tag{12}$$

The local observation likelihood in (10) can also be factorized into intra-group local factor and inter-group local factor:

$$\begin{aligned} \prod_{k \in \mathcal{K}_i^{\mathrm{t}} \cup \{k_1^j\}} p\left(\mathtt{L}_{k-1,k} \mid \mathtt{T}_{k-1}, \mathtt{T}_k\right) \propto & \underbrace{p\left(\mathtt{L}_{k_n^i k_1^j}, \mathtt{L}_{k_1^j} \mid \mathtt{X}_i, \mathtt{X}_j, \mathtt{T}_{k_n^i}\right)}_{\text{inter-group local factor}} \\ & \underbrace{p\left(\mathtt{L}_{k_{\ell_0}^i k_{\ell_1}^i}, \mathtt{L}_{k_{\ell_0}^i} \mid \mathtt{T}_{k_{\ell_1}^i}\right) \prod_{k \in \mathcal{K}_i^{\mathrm{t}} \setminus \{k_{\ell_1}^i\}} p\left(\mathtt{L}_{k-1,k} \mid \mathtt{T}_{k-1}, \mathtt{T}_k\right)}_{\text{intra-group local factor}}, \end{aligned} \tag{13}$$

where $k_{\ell_0}^i = k_{n-N_\ell}^i$ and $k_{\ell_1}^i = k_{n-N_\ell+1}^i$. (12) and (13) are illustrated in the construction of new factors of Fig. 3.

With this formulation, the posterior state set for glocal fusion becomes $\mathcal{X} = \{\mathtt{X}_i, \mathtt{X}_j, \{\mathtt{T}_k\}_{k \in \mathcal{K}_i^{\mathrm{t}}}\}$, and the posterior distribution $p\left(\mathcal{X} \mid \mathcal{Z}\right)$ is proportional to:

$$\begin{aligned} & \underbrace{p\left(\mathtt{X}_i\right)}_{\text{prior}} \underbrace{p\left(\mathtt{L}_{k_{\ell_0}^i k_{\ell_1}^i}, \mathtt{L}_{k_{\ell_0}^i} \mid \mathtt{T}_{k_{\ell_1}^i}\right)}_{1^{\text{st}}\text{ intra-group local}} \underbrace{p\left(\mathtt{L}_{k_n^i k_1^j}, \mathtt{L}_{k_1^j} \mid \mathtt{X}_i, \mathtt{X}_j, \mathtt{T}_{k_n^i}\right)}_{\text{inter-group local}} \\ & \underbrace{\prod_{k \in \mathcal{K}_i^{\mathrm{t}} \setminus \{k_{\ell_1}^i\}} p\left(\mathtt{L}_{k-1,k} \mid \mathtt{T}_{k-1}, \mathtt{T}_k\right)}_{\text{intra-group local}} \underbrace{\prod_{k \in \mathcal{K}_i^{\mathrm{t}}} p\left(\mathtt{G}_k \mid \mathtt{X}_i, \mathtt{T}_k\right)}_{\text{global}} \\ & \underbrace{\prod_{k \in \mathcal{K}_i^{\mathrm{h}}} p\left(\mathtt{G}_k, \mathtt{L}_k \mid \mathtt{X}_i\right) \prod_{k \in \mathcal{K}_j} p\left(\mathtt{G}_k, \mathtt{L}_k \mid \mathtt{X}_j\right)}_{\text{global bundle}}. \end{aligned} \tag{14}$$

MAP estimation is equivalent to minimizing the negative log-posterior. Under the assumption that each likelihood term follows a zero-mean Gaussian distribution, minimizing the negative log-posterior corresponds to solving a weighted least-squares problem. The detailed formulation of each term is provided in Section 5.3. The final factor graph resulting from the proposed glocal fusion approach is shown in Fig. 3(c).

We summarize the advantages of the proposed glocal fusion approach over conventional fusion methods (see Fig. 1) from the perspective of factor graph structure:

- **Global and local consistency preservation:** Conventional fusion directly connects global factors—often noisy and prone to outliers—to local states, risking degradation of the strong local consistency maintained by local SLAM. In contrast, the proposed approach treats locally consistent segments (glocal groups) as fixed random variables, minimizing the direct influence of global observations.
  Moreover, the roles of glocal and local states are clearly separated. Glocal states are optimized to correct global inconsistency without affecting local consistency, while local states serve as smooth transitions between discretely updated glocal states, preserving consistency at group boundaries. This structure avoids the typical trade-off between global and local consistency, ensuring high levels of both.
- **Reduced sensitivity to covariance modeling:** When heterogeneous factors influence the same state, the system becomes highly sensitive to relative covariance models. Since local SLAM covariance depends on environmental observability and is difficult to estimate accurately online, such sensitivity can degrade performance. The pro-

[1] Fixed random variables are not optimized but are assumed to follow a known distribution with predefined covariance.

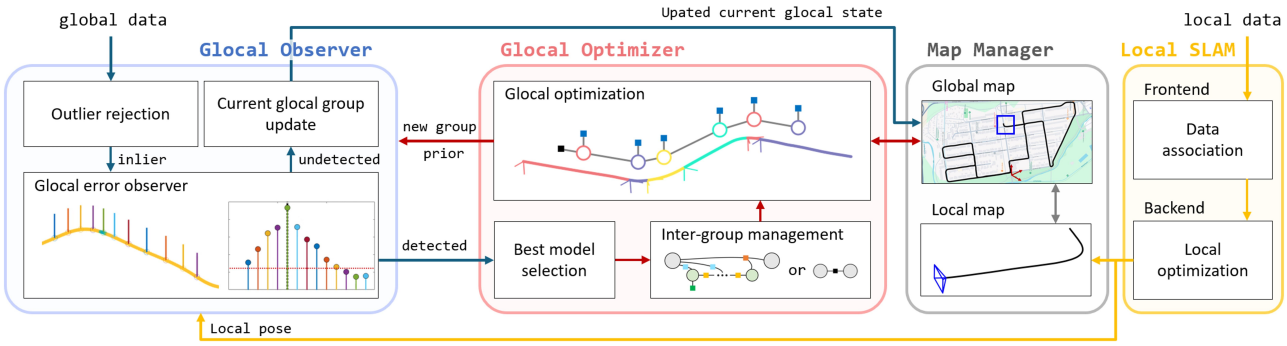

Fig. 4. Framework overview of the proposed glocal navigation system. The system consists of four modules: Local SLAM, Map Manager, and two core modules of the glocal fusion approach: Glocal Observer and Glocal Optimizer.

posed approach reduces this issue by minimizing the structural overlap between heterogeneous factors, thereby improving robustness to covariance mismatch.

- **Robustness via global bundle factor:** Rather than handling global factors individually, they are aggregated into a single bundle factor connected only to the glocal state. This design mitigates the influence of outliers on local states and enables more tractable outlier detection through group-level consistency checks.
- **Optimization efficiency via reduced variables:** By introducing glocal states and modeling most local states as fixed random variables, the proposed approach significantly reduces the number of optimization variables. This enables more efficient computation and allows refinement over longer trajectories within a given window size.

5. Glocal Navigation System

This section describes the proposed glocal navigation system, which builds upon the glocal fusion approach. The overall architecture is first introduced, followed by detailed descriptions of its two main modules: the glocal observer and the glocal optimizer.

5.1. System Overview

The glocal navigation system is systematically designed to apply the glocal fusion approach in an online setting, as shown in Fig. 4. To utilize the proposed approach, we introduce the glocal observer module, which evaluates the trajectory estimated by the local SLAM and detects glocal errors which is drift along the gauge freedom directions. Upon detection, the glocal optimizer is triggered to jointly optimize glocal states and local states between glocal groups.

The system flow is outlined in Fig. 4. Upon receiving local data, the system executes local SLAM corresponding to the sensor modality. The resulting local poses and maps are stored in the map manager. These local poses are included in the current glocal group under evaluation by the glocal observer in order to determine whether they maintain local consistency and whether glocal error has occurred.

Global data typically arrives less frequently than local data. When new global data is received, it activates the glocal observer, which first rejects outliers using the current glocal state. Then, based on the inlier global data, it assesses the presence of glocal error. If no error is detected, only the glocal state is updated, and the global map is refined accordingly. If a glocal error is detected, the observer identifies the segment where the error occurred and activates the glocal optimizer.

In the optimizer, the best model selection identifies the most suitable local states—those acting as bridges between adjacent glocal groups—to minimize optimization error. The inter-group management then determines whether the inter-group local states should be marginalized or reinstated within the current optimization window. Using the resulting graph model, the optimizer refines the associated states. The updated states are passed to the global map, while the glocal observer receives the latest glocal state and prior information to initialize the next evaluation cycle.

5.2. Glocal Observer

The glocal observer module evaluates glocal error by leveraging global observations associated with local poses in the current glocal group. This subsection first defines the glocal error and then introduces the observer model designed to detect such error using both global and local observations. While the specific formulation may vary depending on the types of global observations and the glocal state based on the local SLAM, the derivation process is generally applicable to other combinations. In this subsection, the observer formulation is presented under the case where both the glocal state and global observations lie on SE(3).

5.2.1. Glocal Error

To detect gauge freedom drift within a glocal group, the glocal error is introduced as a parameterization that explicitly expresses the drift. Let the currently evaluated i−th glocal group and its glocal state be denoted by $\mathcal{K}_i$ and $\mathtt{T}_{ew_i}$, respectively. Due to the characteristics of local SLAM, if glocal error arises within $\mathcal{K}_i$, the group can be partitioned into two locally consistent segments separated by the drift. The segment preceding the drift remains correctable by $\mathtt{T}_{ew_i}$, while the segment following the drift is treated as a new glocal group $\mathcal{K}_j$ and corrected using a distinct glocal state $\mathtt{T}_{ew_j}$:

$$\mathtt{T}^*_{ek} = \begin{cases} \mathcal{F}\left(\mathtt{T}_{ew_i}\delta\mathtt{T}_{ew_i}, \tilde{\mathtt{T}}^{\ell}_{w_ik}\right)\mathrm{Exp}\left(\delta\mathsf{n}^{\ell}_k\right), & \left(k \in \mathcal{K}^{\mathrm{h}}_i\right) \\ \mathcal{F}\left(\mathtt{T}_{ew_i}\delta\mathtt{T}_{ew_i}, \mathtt{T}_{w_ik}\delta\mathtt{T}_{w_ik}\right), & \left(k \in \mathcal{K}^{\mathrm{t}}_i\right), \\ \mathcal{F}\left(\mathtt{T}_{ew_j}\delta\mathtt{T}_{ew_j}, \tilde{\mathtt{T}}^{\ell}_{w_jk}\right)\mathrm{Exp}\left(\delta\mathsf{n}^{\ell}_k\right), & \left(k \in \mathcal{K}_j\right) \end{cases} \tag{15}$$

where each $\delta\mathtt{T}$ denotes the state update toward the true value.

The new glocal state $\mathtt{T}_{ew_j}$ can be expressed as:

$$\mathtt{T}_{ew_j}\delta\mathtt{T}_{ew_j} = \mathtt{T}_{ew_i}\delta\mathtt{T}_{ew_i}\mathtt{T}_{w_iw_j}\delta\mathtt{T}_{w_iw_j}. \tag{16}$$

Substituting (16) into the update model for $\mathcal{K}_j$ (15) yields:

$$\begin{aligned} &\mathcal{F}\left(\mathtt{T}_{ew_j}\delta\mathtt{T}_{ew_j}, \tilde{\mathtt{T}}^{\ell}_{w_jk}\right) = \mathtt{T}_{ew_j}\delta\mathtt{T}_{ew_j}\tilde{\mathtt{T}}^{\ell}_{w_jk} \\ &= \mathtt{T}_{ew_i}\delta\mathtt{T}_{ew_i}\mathtt{T}_{w_iw_j}\delta\mathtt{T}_{w_iw_j}\mathtt{T}^{-1}_{w_iw_j}\tilde{\mathtt{T}}^{\ell}_{w_ik} = \mathtt{T}_{ew_i}\delta\mathtt{T}_{ew_i}\delta\mathtt{T}^{w_i}_{w_j}\ \tilde{\mathtt{T}}^{\ell}_{w_ik}, \end{aligned} \tag{17}$$

where $\delta\mathtt{T}^{w_i}_{w_j} = \mathtt{T}_{w_iw_j}\delta\mathtt{T}_{w_iw_j}\mathtt{T}^{-1}_{w_iw_j}$. Here, the glocal error $\delta\mathtt{T}^{w_i}_{w_j}$ represents the gauge freedom drift occurring at the $\{w_j\}$, expressed in the coordinate frame $\{w_i\}$.

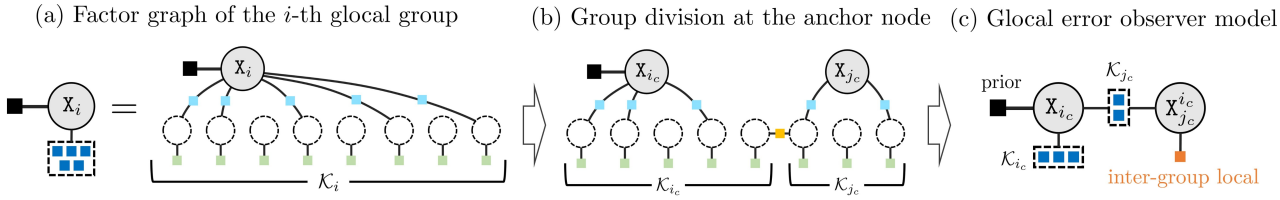

Fig. 5. Glocal error observer. The outline for deriving the glocal error detection model with respect to a specific anchor node within the current glocal group.

5.2.2. Glocal Error Observer Using a Single Anchor Node

This subsection describes the glocal error observer, which detects whether a glocal error has occurred within the current group $\mathcal{K}_i = \{k^i_1, \cdots, k^i_n\}$. To evaluate the presence of glocal error at an anchor node k^{j_c}, the group is partitioned into two segments: $\mathcal{K}_{j_c} = \{k^{j_c}, \cdots, k^i_n\}$ and $\mathcal{K}_{i_c} = \mathcal{K}_i \setminus \mathcal{K}_{j_c}$. The anchor node k^{j_c} serves as a candidate split point for detecting glocal error, and $\mathcal{K}_{j_c}$ corresponds to the newly formed segment starting from the anchor.

During the observer phase, the tail subset $\mathcal{K}^{\mathrm{t}}_i$ is initially set to empty to enable rapid glocal error detection. This subset is later redefined during the best model selection phase in the glocal optimizer, where N_ℓ appropriate local poses are selected for joint optimization with the glocal state. Under this setting, the state update model can be

expressed as: $\mathtt{T}^*_{ek} =$

$$\begin{cases} \mathcal{F}\left(\mathtt{T}_{ew_{i_c}}\delta\mathtt{T}_{ew_{i_c}}, \tilde{\mathtt{T}}^{\ell}_{w_{i_c}k}\right)\mathrm{Exp}\left(\mathsf{n}^{\ell}_{k}\right), & (k \in \mathcal{K}_{i_c}) \\ \mathcal{F}\left(\mathtt{T}_{ew_{i_c}}\delta\mathtt{T}_{ew_{i_c}}\delta\mathtt{T}^{w_{i_c}}_{w_{j_c}}, \tilde{\mathtt{T}}^{\ell}_{w_{i_c}k}\right)\mathrm{Exp}\left(\mathsf{n}^{\ell}_{k}\right), & (k \in \mathcal{K}_{j_c}) \end{cases}. \tag{18}$$

Fig. 5 illustrates the construction of the glocal error detection model within the current glocal group. The factor graph of the i-th glocal group is shown in Fig. 5(a). To evaluate the occurrence of glocal error at a given anchor node, the group is partitioned into $\mathcal{K}_{i_c}$ and $\mathcal{K}_{j_c}$, as depicted in Fig. 5(b). The new glocal state $\mathtt{X}_{j_c}$ is then expressed using the current glocal state and the glocal error. After rearranging the factor graph accordingly, the result is shown in Fig. 5(c). The glocal error observer employs three types of observer models: prior, global bundle, and inter-group local observer models.

Global bundle observer model: The global observation model is defined as $\mathtt{T}^*_{ek} = \tilde{\mathtt{T}}^{g}_{ek}\,\mathrm{Exp}(\mathsf{n}^{g}_{k})$, where $\mathsf{n}^{g}_{k} \sim \mathcal{N}(\mathbf{0}, \Sigma^{g}_{k})$. Combining the global observation model with the state update model in (18), the update $\delta\mathtt{T}$ is expressed in the tangent space. By applying the adjoint map (1), the equation is rearranged as: $\tilde{\mathtt{T}}^{g}_{ek}\,\mathrm{Exp}(\mathsf{n}^{\mathrm{gb}}_{k}) =$

$$\begin{cases} \mathtt{T}_{ew_{i_c}}\tilde{\mathtt{T}}^{\ell}_{w_{i_c}k}\mathrm{Exp}\left(\mathrm{Ad}^{-1}_{\ell_k}\,\delta\boldsymbol{\xi}_{ew_{i_c}}\right), & (k \in \mathcal{K}_{i_c}) \\ \mathtt{T}_{ew_{i_c}}\tilde{\mathtt{T}}^{\ell}_{w_{i_c}k}\mathrm{Exp}\left(\mathrm{Ad}^{-1}_{\ell_k}\left(\delta\boldsymbol{\xi}_{ew_{i_c}} + \delta\boldsymbol{\xi}^{w_{i_c}}_{w_{j_c}}\right)\right), & (k \in \mathcal{K}_{j_c}) \end{cases}, \tag{19}$$

where $\mathsf{n}^{\mathrm{gb}}_{k} = \mathsf{n}^{g}_{k} - \mathsf{n}^{\ell}_{k}$, $\mathrm{Ad}^{-1}_{\ell_k} = \mathrm{Ad}(\tilde{\mathtt{T}}^{\ell}_{w_{i_c}k})^{-1}$. After moving $\tilde{\mathtt{T}}^{g}_{ek}$ to the right-hand side and defining the residual $\boldsymbol{\gamma}^{\mathrm{gb}}_{k}$, the equation is rearranged using BCH formula (2) as follows:

$$\begin{aligned} \boldsymbol{\gamma}^{\mathrm{gb}}_{k} &= \mathrm{Log}\left(\tilde{\mathtt{T}}^{g\,-1}_{ek}\,\mathtt{T}_{ew_{i_c}}\tilde{\mathtt{T}}^{\ell}_{w_{i_c}k}\right) \\ &= \begin{cases} \mathtt{J}^{\mathrm{gb}}_{k}\;\delta\boldsymbol{\xi}_{ew_{i_c}} + \mathsf{n}^{\mathrm{gb}}_{k}, & (k \in \mathcal{K}_{i_c}) \\ \mathtt{J}^{\mathrm{gb}}_{k}\left(\delta\boldsymbol{\xi}_{ew_{i_c}} + \delta\boldsymbol{\xi}^{w_{i_c}}_{w_{j_c}}\right) + \mathsf{n}^{\mathrm{gb}}_{k}, & (k \in \mathcal{K}_{j_c}) \end{cases}, \end{aligned} \tag{20}$$

where $\mathtt{J}^{\mathrm{gb}}_{k} = -\mathtt{J}^{-1}_{\mathtt{r}}(\boldsymbol{\gamma}^{\mathrm{gb}}_{k})\mathrm{Ad}^{-1}_{\ell_k}$, $\mathsf{n}^{\mathrm{gb}}_{k} \sim \mathcal{N}(\mathbf{0}, \Sigma^{\mathrm{gb}}_{k} = \Sigma^{g}_{k} + \Sigma^{\ell})$.

Finally, based on (20), the global bundle observer model can be constructed with respect to the full state update as:

$$\begin{aligned} &\bar{\mathtt{H}}^{\mathrm{gb}} = \begin{bmatrix} \mathtt{H}^{\mathrm{gb}}_{\mathrm{fw}} + \mathtt{H}^{\mathrm{gb}}_{\mathrm{bw}} & \mathtt{H}^{\mathrm{gb}}_{\mathrm{bw}} \\ \mathtt{H}^{\mathrm{gb}}_{\mathrm{bw}} & \mathtt{H}^{\mathrm{gb}}_{\mathrm{bw}} \end{bmatrix}, \quad \bar{\mathbf{g}}^{\mathrm{gb}} = \begin{bmatrix} \mathbf{g}^{\mathrm{gb}}_{\mathrm{fw}} + \mathbf{g}^{\mathrm{gb}}_{\mathrm{bw}} \\ \mathbf{g}^{\mathrm{gb}}_{\mathrm{bw}} \end{bmatrix}, \\ &\text{where} \quad \mathtt{H}^{\mathrm{gb}}_{\square} = \sum_{k \in \mathcal{K}_{\square}} \mathtt{H}^{\mathrm{gb}}_{k}, \qquad \mathbf{g}^{\mathrm{gb}}_{\square} = \sum_{k \in \mathcal{K}_{\square}} \mathbf{g}^{\mathrm{gb}}_{k}, \end{aligned} \tag{21}$$

$\mathtt{H}^{\mathrm{gb}}_{k} = \mathtt{J}^{\mathrm{gb}\top}_{k}\Sigma^{\mathrm{gb}\,-1}_{k}\mathtt{J}^{\mathrm{gb}}_{k}$, and $\mathbf{g}^{\mathrm{gb}}_{k} = \mathtt{J}^{\mathrm{gb}\top}_{k}\Sigma^{\mathrm{gb}\,-1}_{k}\boldsymbol{\gamma}^{\mathrm{gb}}_{k}$. Here, $\bar{\mathtt{H}}^{\mathrm{gb}}$ and $\bar{\mathbf{g}}^{\mathrm{gb}}$ denote the information matrix and gradient, respectively, for the full state update $[\delta\boldsymbol{\xi}^{\top}_{ew_{i_c}},\, \delta\boldsymbol{\xi}^{w_{i_c}\top}_{w_{j_c}}]^{\top}$ in the global bundle observer model. The subscripts fw and bw on $(\cdot)^{\mathrm{gb}}$ denote the forward set $\mathcal{K}_{i_c}$ and the backward set $\mathcal{K}_{j_c}$, respectively.

Inter-group local observer model: As shown in Fig. 5(b and c), the inter-group local observer model is derived from a pair of local poses belonging to different glocal groups ($k_i \in \mathcal{K}_{i_c}$ and $k_j \in \mathcal{K}_{j_c}$). The inter-group local observation ${}^{e}\tilde{\mathtt{T}}^{\ell}_{k_ik_j}$ is defined as the relative pose between these local poses when both belong to the same glocal group without any glocal error:

$$\begin{aligned} &{}^{e}\mathtt{T}^{*}_{k_ik_j} = {}^{e}\tilde{\mathtt{T}}^{\ell}_{k_ik_j}\mathrm{Exp}\left(\mathsf{n}^{\ell}_{k_ik_j}\right), \;\; \mathsf{n}^{\ell}_{k_ik_j} \sim \mathcal{N}\left(\mathbf{0}, \Sigma^{\ell}_{\mathrm{rel}}\right), \\ &\text{where } {}^{e}\tilde{\mathtt{T}}^{\ell}_{k_ik_j} = \mathcal{F}\left(\mathtt{T}_{ew_{i_c}}\delta\mathtt{T}_{ew_{i_c}}, \tilde{\mathtt{T}}^{\ell}_{w_{i_c}k_i}\right)^{-1} \\ &\qquad\qquad \mathcal{F}\left(\mathtt{T}_{ew_{i_c}}\delta\mathtt{T}_{ew_{i_c}}, \tilde{\mathtt{T}}^{\ell}_{w_{i_c}k_j}\right) = \tilde{\mathtt{T}}^{\ell\,-1}_{w_{i_c}k_i}\tilde{\mathtt{T}}^{\ell}_{w_{i_c}k_j}. \end{aligned} \tag{22}$$

where ${}^{e}\tilde{\mathtt{T}}^{\ell}_{k_ik_j}$ denotes the relative pose expressed in the global world frame $\{e\}$, and $\Sigma^{\ell}_{\mathrm{rel}}$ is a system parameter representing the covariance associated with the relative pose estimation from local SLAM. The state update model of ${}^{e}\mathtt{T}^{*}_{k_ik_j}$, following the update equation (18), can be written as:

$${}^{e}\mathtt{T}^{*}_{k_ik_j} = \mathrm{Exp}\left(-\mathsf{n}^{\ell}_{k_i}\right)\tilde{\mathtt{T}}^{\ell\,-1}_{w_{i_c}k_i}\,\delta\mathtt{T}^{w_{i_c}}_{w_{j_c}}\,\tilde{\mathtt{T}}^{\ell}_{w_{i_c}k_j}\mathrm{Exp}\left(\mathsf{n}^{\ell}_{k_j}\right). \tag{23}$$

Following the same derivation procedure as for the global bundle observer model (19) and (20), the residual of

the inter-group local observer model $\boldsymbol{\gamma}^{\text{inter}}_{k_i k_j}$ can be computed as follows:

$$\boldsymbol{\gamma}^{\text{inter}}_{k_i k_j} = \mathbf{0} = \mathtt{J}^{\text{inter}}_{k_i k_j}\, \delta\boldsymbol{\xi}^{w_{ic}}_{w_{jc}} + \mathsf{n}^{\text{inter}}_{k_i k_j}, \tag{24}$$
$$\text{where } \mathtt{J}^{\text{inter}}_{k_i k_j} = -\mathrm{Ad}^{-1}_{\ell_{k_j}},\ \ \mathsf{n}^{\text{inter}}_{k_i k_j} = \mathsf{n}^{\ell}_{k_i k_j} - \mathsf{n}^{\ell}_{k_j} + \mathrm{Ad}^{-1}_{\ell_{k_i k_j}} \mathsf{n}^{\ell}_{k_i},$$

$\mathrm{Ad}^{-1}_{\ell_{k_i k_j}} = \mathrm{Ad}(\tilde{\mathtt{T}}^{\ell\ -1}_{w_{ic} k_i} \tilde{\mathtt{T}}^{\ell}_{w_{ic} k_j})^{-1}$. Since the glocal error observer detects how much glocal error correction is needed under the condition that no glocal error initially exists, the initial value of the glocal error is always set as the identity matrix. Consequently, the residual $\boldsymbol{\gamma}^{\text{inter}}_{k_i k_j}$ becomes a zero vector.

The covariance models for noise terms $\mathsf{n}^{\ell}_{k_i k_j}$ and $\mathsf{n}^{\ell}_{k_j}$ follow $\Sigma^{\ell}_{\text{rel}}$ and Σ^{ℓ}, respectively. However, the covariance for $\mathsf{n}^{\ell}_{k_i}$, denoted by $\Sigma^{\ell}_{k_i}$, does not directly follow Σ^{ℓ}. Instead, it simply compensates for the N_ℓ local states, which are initially empty during the observer phase. The final noise model $\mathsf{n}^{\text{inter}}_{k_i k_j}$ is summarized as follows:

$$\mathsf{n}^{\text{inter}}_{k_i k_j} \sim \mathcal{N}\left(\mathbf{0}, \Sigma^{\text{inter}}_{k_i k_j} = \Sigma^{\ell}_{\text{rel}} + \Sigma^{\ell} + \mathrm{Ad}^{-1}_{\ell_{k_i k_j}} \Sigma^{\ell}_{k_i} \mathrm{Ad}^{-\top}_{\ell_{k_i k_j}}\right),$$
$$\text{where } \Sigma^{\ell}_{k_i} = \mathrm{Ad}^{-1}_{\ell_{k_0 k_i}} \Sigma^{\ell} \mathrm{Ad}^{-\top}_{\ell_{k_0 k_i}} + \sum_{k=k_0+1}^{k_i} \mathrm{Ad}^{-1}_{\ell_{k k_i}} \Sigma^{\ell}_{\text{rel}} \mathrm{Ad}^{-\top}_{\ell_{k k_i}}, \tag{25}$$

where $k_0 = \max\left(k_1^i, k_i - N_\ell\right)$. Lastly, utilizing (24) and (25), we obtain the information matrix and gradient for the inter-group local observer model.

$$\overline{\mathtt{H}}^{\text{inter}} = \begin{bmatrix} 0 & 0 \\ 0 & \mathtt{H}^{\text{inter}} \end{bmatrix}, \quad \overline{\mathbf{g}}^{\text{inter}} = \mathbf{0}, \tag{26}$$
$$\text{where} \quad \mathtt{H}^{\text{inter}} = \mathtt{J}^{\text{inter}\ \top}_{k_i k_j}\ \Sigma^{\text{inter}\ -1}_{k_i k_j}\ \mathtt{J}^{\text{inter}}_{k_i k_j}.$$

Glocal error observer: Based on the previously derived global and local observation models, a module for detecting glocal errors at a given anchor node is presented.

Initially, since no glocal error has been detected within the current group, the system employs the factor graph shown in Fig. 5(a). The resulting optimized glocal state is denoted as $\mathtt{T}_{ew_i}$. Accordingly, the initial state of the glocal error detection model in Fig. 5(c) is set to $\mathtt{T}_{ew_{ic}} = \mathtt{T}_{ew_i}$ and $\delta\mathtt{T}^{w_{ic}}_{w_{jc}} = \mathtt{I}$, corresponding to no glocal error. Using this initialization, the global bundle and inter-group local observer models from (21) and (26), along with the prior factor for the current group ($\mathtt{H}^{\text{p}}$ and $\mathbf{g}^{\text{p}}$) computed by the previous glocal optimization module are combined as:

$$\overline{\mathtt{H}} = \begin{bmatrix} \mathtt{H}_{ew_i} & \mathtt{H}^{\text{gb}}_{\text{bw}} \\ \mathtt{H}^{\text{gb}}_{\text{bw}} & \mathtt{H}^{\text{gb}}_{\text{bw}} + \mathtt{H}^{\text{inter}} \end{bmatrix}, \quad \overline{\mathbf{g}} = \begin{bmatrix} \mathbf{g}_{ew_i} \\ \mathbf{g}^{\text{gb}}_{\text{bw}} \end{bmatrix}, \tag{27}$$
$$\text{where } \mathtt{H}_{ew_i} = \mathtt{H}^{\text{p}} + \mathtt{H}^{\text{gb}}_{\text{fw}} + \mathtt{H}^{\text{gb}}_{\text{bw}},\ \ \mathbf{g}_{ew_i} = \mathbf{g}^{\text{p}} + \mathbf{g}^{\text{gb}}_{\text{fw}} + \mathbf{g}^{\text{gb}}_{\text{bw}}.$$

Here, $\mathtt{H}_{ew_i}$ and $\mathbf{g}_{ew_i}$ denote the information matrix and gradient of the current glocal state $\mathtt{T}_{ew_i}$, respectively.

Since we aim to detect only the glocal error independent of the current glocal state error, the Schur complement is applied to leave only the glocal error:

$$\mathtt{H}_{\text{sc}} = \mathtt{H}^{\text{inter}} + \mathtt{H}^{\text{gb}}_{\text{bw}} - \mathtt{H}^{\text{gb}}_{\text{bw}}\ \mathtt{H}^{-1}_{ew_i}\ \mathtt{H}^{\text{gb}}_{\text{bw}}, \tag{28}$$
$$\mathbf{g}_{\text{sc}} = \quad \mathbf{g}^{\text{gb}}_{\text{bw}} - \mathtt{H}^{\text{gb}}_{\text{bw}}\ \mathtt{H}^{-1}_{ew_i}\ \mathbf{g}_{ew_i}. \tag{29}$$

The information matrix $\mathtt{H}_{\text{sc}}$ represents the degree to which the glocal error at the anchor node is observable from the collected observations, while $\mathbf{g}_{\text{sc}}$ indicates the update direction for the glocal error derived from these observations.

The glocal error at the anchor node is expressed as

$$\delta\boldsymbol{\xi}^{w_{ic}}_{w_{jc}} = \mathtt{H}^{-1}_{\text{sc}}\ \mathbf{g}_{\text{sc}} \sim \mathcal{N}\left(\mathbf{0},\ \Sigma_{\text{sc}} = \mathtt{H}^{-1}_{\text{sc}}\right). \tag{30}$$

The Mahalanobis distance for the error can be computed as:

$$d_{\text{m}} = \delta\boldsymbol{\xi}^{w_{ic}\top}_{w_{jc}} \mathtt{H}_{\text{sc}}\ \delta\boldsymbol{\xi}^{w_{ic}}_{w_{jc}} = \mathbf{g}^{\top}_{\text{sc}}\ \mathtt{H}^{-1}_{\text{sc}}\ \mathbf{g}_{\text{sc}} \sim \chi^2_6\,. \tag{31}$$

Since d_{m} follows a chi-squared distribution with the dimension of the glocal error, a glocal error is detected if d_{m} exceeds the inverse cumulative threshold $\mathrm{CDF}^{-1}_{\chi^2_6}(\beta)$. Intuitively, a large d_{m} indicates a high probability that a glocal error has occurred at the anchor node, necessitating the creation of a new glocal group and state. Conversely, a small d_{m} implies that the current glocal state sufficiently explains the observation residuals, and no additional state is required.

5.2.3. Glocal Error Observer Using Multiple Anchor Nodes

In Section 5.2.2, a method for detecting glocal errors from a single anchor node is described. However, since the precise location where the glocal error occurs within the current group is unknown, this subsection introduces a glocal observer module that employs multiple anchor nodes to identify the error position. The module is systematically designed for online operation. Algorithm 1 outlines the complete procedure of the glocal observer module. The proposed algorithm is applicable regardless of the types of glocal state and global observations, requiring only the internal function definitions to be adapted for specific cases.

The glocal observer module initializes the current glocal group state using the state and prior factor obtained from the glocal optimization. When local data are received, they are appended to the set of local poses $\tilde{\mathbf{L}}$ in the current glocal group. Upon receiving global data, the information matrix and gradient at time k are computed by the `globalObsModel`() function, based on (20) and (21), using the local pose closest in time to the incoming global data $\tilde{\mathbf{G}}_k$. The resulting information matrix and gradient are then added to the glocal state's existing information matrix and gradient. Whether the glocal state requires an update is determined using the Mahalanobis distance, $\mathbf{g}_{ew_i}^{\top}\mathbf{H}_{ew_i}^{-1}\mathbf{g}_{ew_i}$. If an update is required, the `updateGlocalState`() function optimizes the glocal state and updates the relevant internal variables accordingly.

Subsequently, the module updates and evaluates each anchor node within $\mathcal{A}_{\mathcal{J}_c}$. To evaluate an anchor node, the associated information matrix $\mathbf{H}_{\text{sc}}$ and gradient $\mathbf{g}_{\text{sc}}$ of the glocal error are required. The information and gradient for all backward sets of anchor nodes are accumulated into $\mathbf{H}_k^{\text{gb}}$ and $\mathbf{g}_k^{\text{gb}}$, followed by the computation of $\mathbf{H}_{\text{sc}}$ and $\mathbf{g}_{\text{sc}}$ via (28). For anchor nodes with sufficient accumulated observations, the Mahalanobis distance is computed. The reliability of each anchor node is then evaluated using $\mathbf{H}_{\text{sc}} - \mathbf{H}^{\text{inter}}$, which removes the offset introduced by $\mathbf{H}^{\text{inter}}$ and emphasizes the confidence derived solely from observation data. This evaluation prevents the unnecessary creation of new glocal groups due to spurious error detections caused by insufficient observations. An anchor node is regarded as reliable if the log-determinant values of the scale and $\mathrm{SE}(3)$ components of the information matrix are greater than ζ_s and ζ_{T}, respectively.

If the computed d_{m} exceeds a predefined threshold, a glocal error is considered detected. `findBestAnchorNode`() is triggered to select the anchor node with the largest glocal error. The anchor node with the maximum d_{m} is identified, and its neighboring anchor node indices, $\underline{k}$ and $\overline{k}$, are determined. The local pose index corresponding to the largest Mahalanobis distance is then selected by $k^{j_c^*} = \text{argmax}_{k^{jc}}\, d_{\text{m}}(k^{j_c})$, where $k^{j_c} \in \{\underline{k}, \cdots, \overline{k}\}$. The selected best anchor node is returned to the glocal optimizer.

To limit computational overhead, the number of anchor nodes is constrained by a maximum value N_{anc}. Upon receiving new global data, a new anchor node is created at the local pose closest in time to the incoming data. If the number of anchor nodes exceeds N_{anc}, anchor node deletion is triggered. Since neighboring anchor nodes often share similar data and detect errors at proximate trajectory points, such redundancy is reduced by deleting nodes according to two criteria: (1) anchor nodes in the first half that have passed through multiple evaluations, and (2) anchor nodes with minimal differences in d_{m} compared to adjacent nodes and low likelihood of glocal error (i.e., low d_{m}).

5.2.4. Outlier Rejection

Since global observations are generally noisy and often contain outliers, this subsection describes the outlier rejection method in the glocal fusion model. Given a global observation $\tilde{\mathbf{T}}_{ek}^{g}$, its residual and Jacobian with respect to the current glocal state can be computed as follows:

$$\begin{aligned} \boldsymbol{\gamma}_k^{\text{g}} &= \text{Log}\left(\tilde{\mathbf{T}}_{ek}^{g\,-1}\,\mathbf{T}_{ew_i}\tilde{\mathbf{T}}_{w_ik}^{\ell}\right) = \mathbf{J}_k^{\text{g}}\delta\boldsymbol{\xi}_{ew_i} + \mathbf{n}_k^{\text{gb}} = \tilde{\mathbf{n}}_k^{\text{g}}, \\ \text{where } \tilde{\mathbf{n}}_k^{\text{g}} &\sim \mathcal{N}\left(\mathbf{0}, \tilde{\Sigma}_k^{\text{g}} = \mathbf{J}_k^{\text{g}}\,\mathbf{H}_{ew_i}^{-1}\mathbf{J}_k^{\text{g}\top} + \Sigma_k^{\text{gb}}\right), \\ \mathbf{J}_k^{\text{g}} &= -\mathbf{J}_{\text{r}}^{-1}\left(\boldsymbol{\gamma}_k^{\text{g}}\right)\text{Ad}_{\ell_k}^{-1}, \quad \delta\boldsymbol{\xi}_{ew_i} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{H}_{ew_i}^{-1}\right). \end{aligned} \tag{32}$$

Here, $\tilde{\mathbf{n}}_k^{\text{g}}$ represents the noise model of the residual $\boldsymbol{\gamma}_k^{\text{g}}$ for outlier detection. The residual is normalized using its covariance, and the Mahalanobis distance is computed for outlier rejection. A global observation is considered an outlier if it satisfies the following condition: $\boldsymbol{\gamma}_k^{\text{g}\top}\,\tilde{\Sigma}_k^{\text{g}^{-1}}\boldsymbol{\gamma}_k^{\text{g}} \geq \zeta_{\text{out}} := \text{CDF}_{\chi_6^2}^{-1}\left(\beta_{\text{out}}\right)$.

Algorithm 1: Glocal Observer

Input: $\mathtt{X}_{ew_i}$, $\mathtt{H}^{\mathsf{p}}$, $\mathbf{g}^{\mathsf{p}}$: data for current glocal group
Output: $\mathtt{A}_{j_c^*}$: best anchor node

```
// Initialize current glocal group
H_{ew_i} ← H^p;   g_{ew_i} ← g^p;
for each incoming data D̃_k do
    if isLocalData(D̃_k) then
        L̃.insert(L̃_k) continue;
    // Process global data
    [H^gb_k, g^gb_k] ← globalObsModel(X_{ew_i}, L̃, G̃_k);
    // Update for glocal state
    H_{ew_i} ← H_{ew_i} + H^gb_k;   g_{ew_i} ← g_{ew_i} + g^gb_k;
    if checkGlocalState(H_{ew_i}, g_{ew_i}) then
        X_{ew_i} ← updateGlocalState();
    // Update and evaluate anchor nodes
    for each A_{j_c} ∈ A_{J_c} do
        A_{j_c}.H^gb_bw ← A_{j_c}.H^gb_bw + H^gb_k;
        A_{j_c}.g^gb_bw ← A_{j_c}.g^gb_bw + g^gb_k;
        schurComplement(A_{j_c});
        if checkEnoughInformation(A_{j_c}) then
            computeMahaDist(A_{j_c});
    // Detect glocal error
    if checkGlocalError(A_{J_c}) then
        A_{j_c^*} ← findBestAnchorNode(A_{J_c});
        return A_{j_c^*};
    // Create or delete anchor node
    if checkNeedNewAnchor() then
        A_{j_c^new} ← initializeNewAnchor();
        A_{J_c}.insert(A_{j_c^new});
        if checkNeedDeleteAnchor() then
            findDeleteAnchorAndDelete(A_{J_c});
```

5.3. Glocal Optimizer

The glocal optimizer module creates a new glocal group based on the glocal error detection results from the glocal observer module and subsequently performs sliding-window optimization. As illustrated in Fig. 4, the optimizer comprises three submodules: best model selection, inter-group management, and glocal optimization.

5.3.1. Best Model Selection

The best model selection submodule reintroduces N_ℓ local states that are not modeled in the glocal observer module, which aimed for fast and efficient glocal error detection. This submodule selects the local state model that minimizes the observation residual. Let $k^{j_c^*}$ denote the local pose index at which a glocal error is identified by the observer. Based on this point, the set of local pose indices eligible to be local states is defined as:

$$\mathcal{K}_{\text{cand}} = \{\max\left(k_1^i, k^{j_c^*}-N_\ell+1\right), \cdots, \min\left(k_n^i, k^{j_c^*}+N_\ell-1\right)\},$$

where k_1^i and k_n^i denote the indices of the first and last local poses in the current glocal group, respectively.

Subsets $\mathcal{K}_i^{\mathrm{tc}}$ containing N_ℓ consecutive elements from the candidate set $\mathcal{K}_{\mathrm{cand}}$ are constructed as candidate local state models. For each candidate, a factor graph is built as illustrated in Fig. 3(c), and optimization is performed. After optimization, residuals are normalized using their respective covariance models, and the sum of squared normalized residuals is computed. The candidate model yielding the minimum sum is selected as the best model $\mathcal{K}_i^{\mathrm{t}}$. In this study, three uniformly distributed candidate models from $\mathcal{K}_{\mathrm{cand}}$ are evaluated to determine the best model.

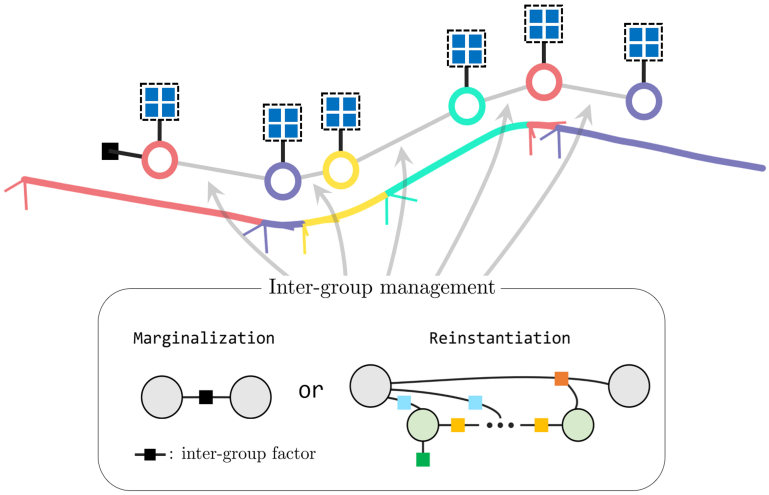

Fig. 6. Glocal optimization framework with inter-group management deciding marginalization or reinstantiation of local states.

5.3.2. Inter-Group Management

The inter-group management determines how local states that connect glocal states within the optimization window are managed. These local states compensate for gauge freedom drift around the boundary between adjacent glocal groups and ensure smooth transitions between discretely updated glocal states, thereby maintaining consistency across group boundaries. When the relative transformation between adjacent glocal states changes significantly, these local states play a critical role. However, after a few glocal optimizations, the relative glocal states typically converge, resulting in only minor changes thereafter. In such cases, keeping all inter-group local states as optimization variables becomes computationally inefficient. Instead, it is preferable to marginalize them and represent their influence as inter-group factors connecting the adjacent glocal states.

As illustrated in Fig. 6, inter-group management determines whether to marginalize or reinstantiate the inter-group local states before glocal optimization. After each glocal optimization, any local states retained as optimization variables are marginalized to construct inter-group factors. The relative glocal states between the associated groups at the time of marginalization are also stored.

Using the stored relative glocal states and inter-group factors, inter-group management determines whether marginalization or reinstantiation is necessary based on two criteria: (1) the deviation between the stored relative glocal state (at the time of marginalization) and the current relative glocal state is evaluated. If the deviation is large, the previously marginalized local states are reinstantiated. (2) The information matrix $\mathbf{H}^{\mathrm{inter}}$ and gradient $\mathbf{g}^{\mathrm{inter}}$ of the inter-group factor are computed using the current connected glocal states. Based on these, the Mahalanobis distance is calculated to determine whether the factor is still valid. A large Mahalanobis distance implies that the current inter-group factor no longer supports the glocal states well, prompting reinstantiation and reoptimization of the local states, followed by factor update.

In addition, the inter-group factor that connects the oldest glocal state within the optimization window is always reinstantiated. This is because once a glocal group moves out of the window, it is no longer updated and is summarized into a prior factor. Therefore, having the most accurate state before marginalization is desirable.

5.3.3. Glocal Optimization

Glocal optimization jointly optimizes the set of variables $\mathcal{X}$ consists of the glocal states within the window and the local states reinstantiated by inter-group management, as shown in the factor graph in Fig. 6.

$$\begin{aligned} \mathcal{X} &:= \left\{ \{\mathtt{X}_i\}_{i\in\mathcal{W}},\ \{\mathtt{T}_{ik}\}_{k\in\mathcal{K}_i^{\mathsf{t}}}^{i\in\mathcal{W}_{\mathcal{I}}} \right\}, \\ \text{where } \mathcal{W} &= \{i_1, ..., i_{N_{\mathsf{ws}}}\} = \mathcal{W}_{\mathcal{I}} \cup \mathcal{W}_{\mathcal{M}} \cup \{i_{N_{\mathsf{ws}}}\}, \\ \mathcal{K}_i^{\mathsf{t}} &= \{k_i^{\mathsf{t}_1}, \cdots, k_i^{\mathsf{t}_e}\}, \quad \mathcal{K}_{i_{N_{\mathsf{ws}}}}^{\mathsf{t}} = \emptyset\ . \end{aligned} \tag{33}$$

Here, $\mathcal{W}$ denotes the set of indices of the glocal states within the optimization window. The sets $\mathcal{W}_{\mathcal{I}}$ and $\mathcal{W}_{\mathcal{M}}$ contain the indices of the preceding groups connected by the reinstantiation and marginalization local factor models, respectively. For notation simplicity, the i-th glocal state $\mathtt{X}_{ew_i}$ and the k-th local pose $\mathtt{T}_{w_ik}$ within the i-th group are denoted by $\mathtt{X}_i$ and $\mathtt{T}_{ik}$, respectively. The optimization cost is constructed as follows:

$$\begin{aligned} \mathcal{X}^* = \underset{\mathcal{X}}{\text{argmin}} \Bigg\{ & \|\boldsymbol{\gamma}_{i_1}^{\mathsf{p}}\|^2_{\Sigma_{i_1}^{\mathsf{p}}} + \sum_{i\in\mathcal{W}} \|\boldsymbol{\gamma}_i^{\mathsf{gb}}\|^2_{\Sigma_i^{\mathsf{gb}}} + \sum_{i\in\mathcal{W}_{\mathcal{M}}} \|\boldsymbol{\gamma}_{i,i+1}^{\mathsf{inter}}\|^2_{\Sigma_{i,i+1}^{\mathsf{inter}}} \\ & \sum_{i\in\mathcal{W}_{\mathcal{I}}} \left\{ \sum_{k\in\mathcal{K}_i^{\mathsf{g}}\cap\mathcal{K}_i^{\mathsf{t}}} \|\boldsymbol{\gamma}_{i,k}^{\mathsf{g}}\|^2_{\Sigma_{i,k}^{\mathsf{g}}} + \sum_{k\in\mathcal{K}_i^{\mathsf{t}}} \|\boldsymbol{\gamma}_{i,i+1,k}^{\ell}\|^2_{\Sigma_{i,i+1,k}^{\ell}} \right\} \Bigg\}, \end{aligned} \tag{34}$$

The superscripts p, gb, inter, g, and ℓ represent the prior, global bundle, inter-group, global, and local factors, respectively. The local factors include both intra-group and inter-group local factors. The subscripts indicate the indices of the states constrained by each factor. The set $\mathcal{K}_i^{\mathsf{g}}$ denotes the indices of local poses in the i-th group that are closest in time to the global observations.

The state update model is defined as follows: $\mathtt{T}_{ek}^* =$

$$\begin{aligned} & \begin{cases} \mathcal{F}\left(\mathtt{X}_i\delta\mathtt{X}_i, \tilde{\mathtt{T}}_{ik}^{\ell}\right)\mathtt{N}_k^{\ell} &= \mathtt{F}_{ik}^{\ell}\mathsf{Exp}\left(\mathtt{J}_{\mathsf{h}}\delta\mathbf{x}_i\right)\mathtt{N}_k^{\ell}, \qquad (k\in\mathcal{K}_i^{\mathsf{h}}) \\ \mathcal{F}\left(\mathtt{X}_i\delta\mathtt{X}_i, \mathtt{T}_{ik}\delta\mathtt{T}_{ik}\right) &= \mathtt{F}_{ik}\mathsf{Exp}\left(\mathtt{J}_{\mathsf{t}}[\delta\mathbf{x}_i^{\top}, \delta\boldsymbol{\xi}_{ik}^{\top}]^{\top}\right), (k\in\mathcal{K}_i^{\mathsf{t}}) \end{cases}, \\ & \text{where } \mathtt{N}_k^{\ell} = \mathsf{Exp}\left(\mathfrak{n}_k^{\ell}\right), \mathtt{F}_{ik}^{\ell} = \mathcal{F}\left(\mathtt{X}_i, \tilde{\mathtt{T}}_{ik}^{\ell}\right), \mathtt{F}_{ik} = \mathcal{F}(\mathtt{X}_i, \mathtt{T}_{ik}). \end{aligned} \tag{35}$$

Here, the Jacobian $\mathtt{J}$ corresponds to the state update model and is derived for each case.

The global observation model is defined as $\mathsf{f}_{\mathsf{g}}\left(\mathtt{T}_{ek}^*\right) = \tilde{\mathtt{G}}_k \oplus \mathfrak{n}_k^g$, where $\mathsf{f}_{\mathsf{g}}(\mathtt{T})$ maps $\in \mathrm{SE}(3)$ to the glocal observation space. $\oplus$ denotes the noise addition operator, which depends on the type of $\mathtt{G}_k$. Using the observation model and (35), the residuals for the global bundle and global factors are defined as:

$$\begin{aligned} & \boldsymbol{\gamma}_i^{\mathsf{gb}} = \sum_{k\in\mathcal{K}_i^{\mathsf{g}}\cap\mathcal{K}_i^{\mathsf{h}}} \mathtt{F}_{ik}^{\ell} \ominus \tilde{\mathtt{G}}_k, \qquad \boldsymbol{\gamma}_{ik}^{\mathsf{g}} = \mathtt{F}_{ik} \ominus \tilde{\mathtt{G}}_k, \\ & \text{where } \mathtt{F}\ominus\mathtt{G} = \begin{cases} \mathsf{Log}\left(\mathtt{G}^{-1}\mathtt{F}\right), & (\mathtt{G}\in\mathrm{SE}(3)) \\ \mathsf{f}_{\mathsf{g}}(\mathtt{F}) - \mathtt{G}, & (\mathtt{G}\in\text{vector space}) \end{cases}. \end{aligned} \tag{36}$$

The local observation model is defined as:

$$\begin{aligned} {}^{e}\mathtt{T}_{k_ik_j}^* &= {}^{e}\tilde{\mathtt{T}}_{k_ik_j}^{\ell}\mathsf{Exp}(\mathfrak{n}_{k_ik_j}^{\ell}), \\ \text{where } {}^{e}\mathtt{T}_{k_ik_j}^* &= \mathtt{T}_{ek_i}^{*\,-1}\,\mathtt{T}_{ek_j}^*, \quad {}^{e}\tilde{\mathtt{T}}_{k_ik_j}^{\ell} = \mathtt{T}_{ek_i}^{\ell\,-1}\,\mathtt{T}_{ek_j}^{\ell}. \end{aligned} \tag{37}$$

Here, the pose $\mathtt{T}_{ek}^*$ is determined by the state update model in (35), and $\mathtt{T}_{ek}^{\ell}$ is defined by $\mathcal{F}(\mathtt{X}_i\delta\mathtt{X}_i, \tilde{\mathtt{T}}_{ik}^{\ell})$. For the inter-group local factor, k_j belongs to the j-th group. In this case, the state update model is defined with respect to group j, whereas the observation remains defined with respect to group i, without considering the group division:

$$\mathtt{T}_{ek_j}^* = \mathtt{F}_{jk_j}^{\ell}\mathsf{Exp}\left(\mathtt{J}_{\mathsf{h}}\delta\mathbf{x}_j\right)\mathtt{N}_{k_j}^{\ell}, \quad \mathtt{T}_{ek_j}^{\ell} = \mathcal{F}\left(\mathtt{X}_i\delta\mathtt{X}_i, \tilde{\mathtt{T}}_{ik}^{\ell}\right). \tag{38}$$

By (37) and (38), the local factor residual $\boldsymbol{\gamma}_{i,i+1,k}^{\ell}$ is given as:

$$\begin{cases} \mathsf{Log}(\mathtt{F}_{ik_j}^{\ell\,-1}\mathtt{F}_{ik_j}), & (k_i\in\mathcal{K}_i^{\mathsf{h}}, k_j\in\mathcal{K}_i^{\mathsf{t}}) \\ \mathsf{Log}(\mathtt{F}_{ik_j}^{\ell\,-1}\mathtt{F}_{ik_i}^{\ell}\mathtt{F}_{ik_i}^{-1}\mathtt{F}_{ik_j}), & (k_i, k_j\in\mathcal{K}_i^{\mathsf{t}}) \\ \mathsf{Log}(\mathtt{F}_{ik_j}^{\ell\,-1}\mathtt{F}_{ik_i}^{\ell}\mathtt{F}_{ik_i}^{-1}\mathtt{F}_{i+1,k_j}^{\ell}), & (k_i\in\mathcal{K}_i^{\mathsf{t}}, k_j\in\mathcal{K}_{i+1}^{\mathsf{h}}) \end{cases}. \tag{39}$$

The first and second cases refer to the intra-group local factors, while the last case corresponds to the inter-group local factor. The intra-group local factors influence only the local states, while the inter-group local factor always affects both the connected glocal states and the local state, regardless of the glocal state representation. In addition, when the glocal state is defined in $\mathrm{SIM}(3)$, a regularization term is added to preserve scale consistency between adjacent groups, preventing the scale from being fully corrupted by noise in global observations. The Jacobian and covariance of each factor follow the derivation in Section 5.2.2 and are omitted here for brevity.

After glocal optimization, two types of marginalization are performed. First, the local states in $\mathcal{W}_{\mathcal{I}}$ are marginalized to form inter-group factors, which are subsequently used in inter-group management. Second, all states except the last glocal state are marginalized to generate a prior factor for the last group. This prior is used in the next glocal observer step (27).

6. Experimental Results

This section presents the experimental results of the proposed glocal fusion system using various combinations of local and global data, and compares its performance with that of conventional fusion systems.

6.1. Experimental Setup

6.1.1. Local and Global Data

This paper proposes a general sensor fusion framework capable of integrating various types of local and global data. To validate the proposed system, experiments are conducted using various local SLAM methods and global observations. For local data, we employ widely used SLAM systems based on different sensor modalities: ORB-SLAM (monocular and stereo versions)[1], VINS-Fusion (stereo version)[10], and A-LOAM (LiDAR)[4]. For clarity, each local SLAM system is referred to by an abbreviation throughout this section: OM for ORB-SLAM (monocular), OS for ORB-SLAM (stereo), VF for VINS-Fusion, and LM for A-LOAM.

To evaluate the fusion system under various levels of global observation noise, synthetic global data are generated using ground-truth poses from the dataset. For each odometry sequence, the ground-truth poses are transformed into a randomly generated global coordinate frame $\{e\}$ to produce $\{\mathtt{T}^{*}_{ek}\}_k$. Noise $\mathsf{n}^g_k \sim \mathcal{N}(\mathbf{0}, \Sigma^g_k)$ is then added to generate global observations. The covariance Σ^g_k is computed as: $\sqrt{\Sigma^g_k} = \mathtt{U}_k \cdot \mathrm{diag}\,(N_{\text{g-lvl}} \cdot \bar{\mathsf{a}}_k) \cdot \mathtt{U}_k^{\top}$, where $\mathtt{U}_k$ is an orthonormal matrix, and $\bar{\mathsf{a}}_k$ is a 6D normalized vector. The parameter $N_{\text{g-lvl}}$ controls the noise level of the standard deviation. For each pose, $\mathtt{U}_k$ and $\bar{\mathsf{a}}_k$ are randomly sampled to generate the corresponding covariance and global observation. The resulting global observations are denoted by $\mathtt{T}(N_{\text{g-lvl}})$ and $\mathsf{p}(N_{\text{g-lvl}})$, indicating pose and position data generated with noise level $N_{\text{g-lvl}}$. For instance, $\mathtt{T}(1)$ refers to a global observation with standard deviations of 1° in rotation and $1\,\mathrm{m}$ in translation.

To this end, we use the KITTI odometry dataset[40], which provides ground-truth data along with LiDAR and stereo inputs, to conduct various ablation experiments and quantitatively compare the proposed system with conventional fusion methods.

6.1.2. Baseline: Conventional Fusion System

This subsection describes the conventional fusion system used as a baseline for evaluating the proposed glocal fusion system. Before optimizing the factor graph structure shown in Fig. 2, conventional fusion systems typically perform coordinate alignment between global and local data to ensure stability. Similarly, the baseline in this paper first estimates a global state that represents the relative transformation between the local world frame $\{w\}$ of the local SLAM system and the global world frame $\{e\}$, referred to as global alignment. To avoid over-parameterization, local SLAM's states are assumed to be accurate, and only the global state is estimated during alignment.

Performing fusion before accurately estimating the global state may lead to poor estimation or convergence to undesirable local minima. Therefore, the system waits until the local observability Gramian (LOG), which quantifies the accumulated information for estimating the global state, becomes sufficiently informative. Specifically, the log-determinant of the LOG must exceed the predefined thresholds ζ_s and ζ_{T} for the scale and $\mathrm{SE}(3)$ components, respectively. When the global observation consists only of positional information, the rotation about the direction of motion becomes unobservable during straight-line trajectories, potentially resulting in incorrect global state estimation. To address this issue, when only global positions are available, the rotation component of the LOG is

extracted via Schur complement, and the log-ratio between its maximum and minimum eigenvalues is evaluated against the threshold γ_{rot}.

Once the LOG is sufficient, the global state estimated from the alignment phase is used to initialize the local state nodes $\mathbf{T}_{ek}$ in the factor graph shown in Fig. 1(a). Then, all local states used during global alignment are jointly optimized using both local and global factors. Subsequently, sliding-window optimization is performed with a fixed window size N_{ws}. Old states that exit the window are marginalized to construct prior factors. For optimization of the conventional fusion, GTSAM[42] version 4.2 is employed with the Levenberg-Marquardt method. In contrast, the proposed glocal optimizer utilizes a custom optimization solver developed by the author, also based on the Levenberg-Marquardt method.

6.1.3. Hyperparameters

The hyperparameters used for the proposed glocal fusion and the conventional fusion systems are summarized in Table 1. The parameter ζ, used to evaluate the reliability of each anchor in the glocal observer and of global alignment in the conventional fusion, is set based on estimation theory, such that it corresponds to an n-sigma confidence level for a given threshold. For instance, ζ_{T} in the glocal observer is determined to correspond to the information level required for bounding rotation and translation errors within 1.5° and $1.5\,\text{m}$, respectively, at the 3-σ confidence level.

The covariance $\Sigma^{\ell}_{\text{rel}}$ of relative poses in local SLAM, a key parameter for optimization, is identically set for both the conventional and glocal fusion systems, and an ablation study is reported in Section 6.2. In conventional fusion, N_{ws} limits the number of local states, whereas in the glocal optimizer, the window size constrains the number of glocal states. The inter-group factors reinstantiated by the glocal fusion system's inter-group management module average between 1.1 and 1.3 (see Table 2). For fairness, N_{ws} in each system is therefore selected to yield a comparable number of states optimized per iteration. Unless otherwise specified for ablation studies, all experiments are conducted under identical parameter settings.

Table 1. Hyperparameters for the proposed glocal fusion and conventional fusion systems.

Glocal Fusion	Param	Value
Glocal Observer		
ㄴ local covariance	Σ^{ℓ}	$\text{Cov}(\sigma^{\ell}_{\omega}=0.05^\circ,\ \sigma^{\ell}_{\nu}=0.05\,\text{m})$
ㄴ detect outlier	ζ_{out}	$\text{CDF}^{-1}_{\chi^2}\,(\beta_{\text{out}}=5\text{-}\sigma)$
ㄴ detect glocal error	ζ_{err}	$\text{CDF}^{-1}_{\chi^2}\,(\beta_{\text{err}}=1\text{-}\sigma)$
ㄴ check Hessian	ζ_s	$\log(\mathbf{H}_s=\frac{2}{\log(1.1)})$
	ζ_{T}	$\log(\det(\mathbf{H}_{\text{T}}=\text{Cov}(\frac{1.5^\circ}{3},\ \frac{1.5\,\text{m}}{3})^{-1}))$
ㄴ # of anchor nodes	N_{anc}	200
Glocal Optimizer		
ㄴ local rel. covar.	$\Sigma^{\ell}_{\text{rel}}$	$\text{Cov}(\sigma^{\ell}_{\omega,\text{rel}}=0.04^\circ, \sigma^{\ell}_{\nu,\text{rel}}=0.04\,\text{m})$
ㄴ window size	N_{ws}	10
ㄴ # of local states	N_{ℓ}	10

Conventional Fusion	Param	Value
Global alignment		
ㄴ check Hessian	ζ_s	$\log(\mathbf{H}_s=\frac{3}{\log(1.1)})$
	ζ_{T}	$\log(\det(\mathbf{H}_{\text{T}}=\text{Cov}(\frac{0.3^\circ}{3},\ \frac{0.3\,\text{m}}{3})^{-1}))$
	γ_{rot}	6
Optimization		
ㄴ local rel. covar.	$\Sigma^{\ell}_{\text{rel}}$	$\text{Cov}(\sigma^{\ell}_{\omega,\text{rel}}=0.04^\circ,\ \sigma^{\ell}_{\nu,\text{rel}}=0.04\text{m})$
ㄴ detect outlier	ζ_{out}	$\text{CDF}^{-1}_{\chi^2}\,(\beta_{\text{out}}=5\text{-}\sigma)$
ㄴ window size	N_{ws}	25

∗ $\text{Cov}(\sigma_{\omega},\sigma_{\nu})=\text{diag}\left([(\frac{\sigma_{\omega}\pi}{180})^2\times[1,1,1],\sigma_{\nu}^2\times[1,1,1]]\right)$.
∗ The 1-σ and 5-σ are $68.27\,\%$ and $99.99994\,\%$, respectively.

6.1.4. Error Metrics

This subsection describes the error metrics used to evaluate the global and local consistency of sensor fusion systems. To evaluate global consistency, the absolute trajectory error (ATE) is used, while the relative trajectory error (RTE) is employed for assessing local consistency. The error terms are defined as follows:

$$\mathrm{ATE} = \left(\frac{1}{l} \sum_{k=1}^{l} \| \mathrm{trans}(\mathbf{T}_{ek}^{*^{-1}} \mathbf{T}_{ek}) \|^2 \right)^{0.5}, \tag{40}$$

$$\mathrm{RTE} = \frac{1}{\mathrm{Card}(\mathcal{I})} \sum_{\Delta \in \mathcal{I}} \frac{100}{\Delta} \left(\frac{1}{l(\Delta)} \sum_{k=1}^{l(\Delta)} \| \mathrm{trans}(\mathbf{E}_k) \|^2 \right)^{0.5},$$

where $\mathbf{E}_k = (\mathbf{T}_{ek}^{*^{-1}} \mathbf{T}_{ek(\Delta)}^{*})^{-1} (\mathbf{T}_{ek} \mathbf{T}_{ek(\Delta)})$. Δ denotes the fixed traveling length interval used for RTE calculation, and the set of all intervals is represented by $\mathcal{I}$. The index $k(\Delta)$ indicates the frame that is at least Δ away from the k-th frame and closest to Δ, while $l(\Delta)$ represents the total number of frames for which RTE can be calculated using Δ. In this experiment, to evaluate estimation accuracy in the local domain, $\mathcal{I}$ is set to $\{10, 20, 30, 40, 50\,\mathrm{m}\}$.

Unless otherwise specified, the unit of ATE is m, and the unit of RTE is m, which can be intuitively interpreted as the average error in meters per 100 meters of traveled distance. All reported results are obtained by averaging over five repeated trials under identical conditions.

6.2. Ablation Study

To demonstrate the role of key modules and hyperparameters in the proposed glocal fusion system, an ablation study is conducted by varying different hyperparameters and evaluating their impact on performance. The performance of each system is evaluated by the average ATE and RTE across all odometry sequences of the KITTI dataset.

6.2.1. Local SLAM Relative Covariance

This section presents an ablation study on the covariance $\Sigma_{\mathrm{rel}}^{\ell}$ of relative poses in local SLAM, which is a key parameter in the optimization of both the glocal and conventional fusion systems. As discussed earlier, computing this covariance online is challenging, and fusion systems therefore rely on pre-defined values. A mismatch between the pre-defined and true covariances can degrade performance, and robust systems are expected to tolerate such discrepancies. To evaluate this robustness, $\Sigma_{\mathrm{rel}}^{\ell}$ is systematically varied over a wide range and the resulting performance of the fusion systems is analyzed.

The covariance $\Sigma_{\mathrm{rel}}^{\ell}$ is parameterized by the rotation standard deviation $\sigma_{\omega,\mathrm{rel}}^{\ell}$ and the translation standard deviation $\sigma_{\nu,\mathrm{rel}}^{\ell}$, which are varied from 0.01° to 0.1° and from $0.01\,\mathrm{m}$ to $0.1\,\mathrm{m}$, respectively. The evaluation employs global pose and position observations with a noise level of 2.5, and local systems based on OM and OS. Specifically, the glocal fusion is tested with $\mathbf{T}, \mathbf{p}(2.5)$-OM using $\mathrm{SIM}(3)$ as the glocal state, and with $\mathbf{T}, \mathbf{p}(2.5)$-OS using $\mathrm{SE}(3)$. Fig. 7 shows the results.

As illustrated in Fig. 7, the proposed glocal fusion system exhibits significantly higher robustness to variations in local relative covariance compared to the conventional system. The glocal fusion consistently preserves both global and local consistency across a broad range of covariance settings. In contrast, the conventional system is sensitive to the covariance, and its performance degrades severely, sometimes even diverging, when $\sigma_{\nu,\mathrm{rel}}^{\ell}$ is under-determined. This robustness stems from the glocal fusion approach minimizing structural overlap between heterogeneous factors in the factor graph (see Section 4.3), thereby reducing sensitivity to covariance mismatch. Furthermore, the performance gap is particularly pronounced in local consistency, as the proposed method groups locally consistent states into glocal groups, which prevents degradation from noisy global factors.

Based on these results, the local relative covariance is set to $\mathrm{Cov}(0.04^\circ, 0.04\,\mathrm{m})$, which yields the best overall performance for both the proposed and baseline systems.

6.2.2. Optimization Window Size

This subsection compares the performance of each system with respect to the optimization window size. For the conventional fusion system, N_{ws} is varied, while for the glocal fusion system, N_{ℓ} is varied. Since $N_{\mathrm{ws}} = 10$ already provides a sufficiently large window for glocal fusion, its performance is instead evaluated with respect to N_{ℓ}, which determines the number of local states optimized across groups. The fusion cases used for evaluation are identical

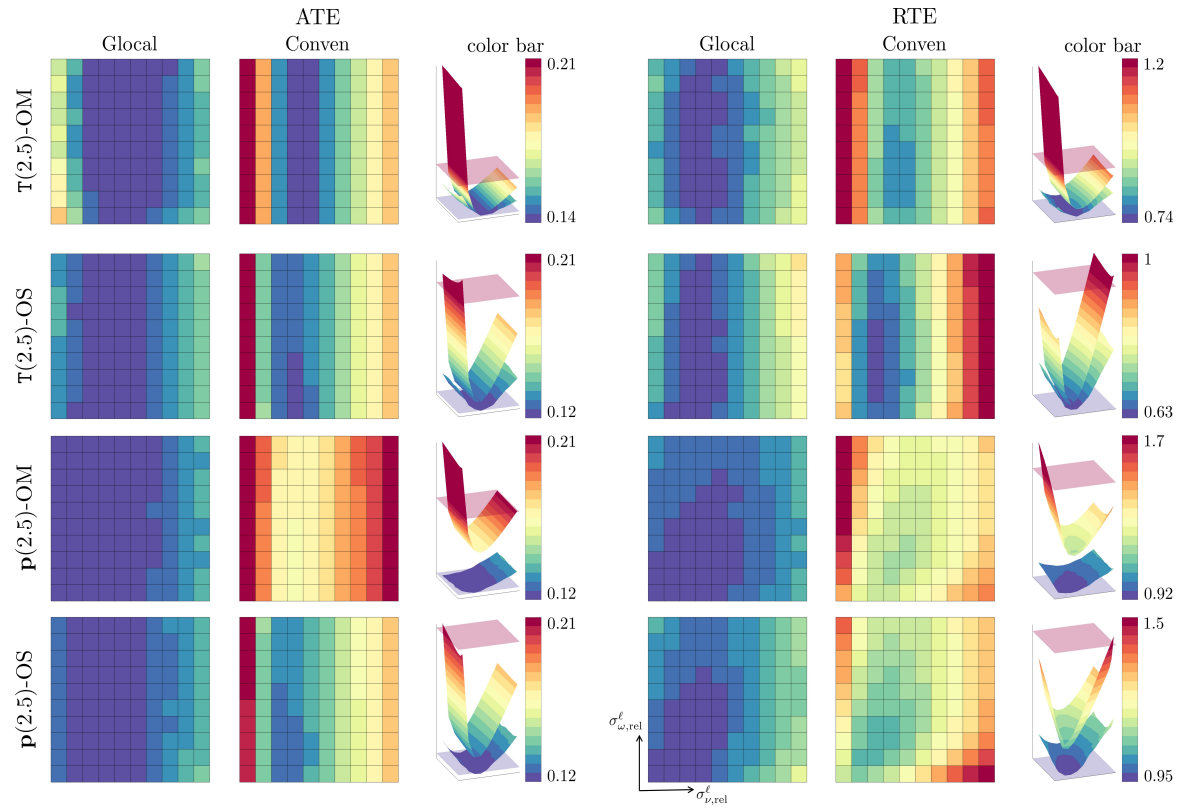

Fig. 7. Performance of glocal and conventional fusion systems under varying local relative covariance in four fusion cases. $\sigma^{\ell}_{\omega,\mathrm{rel}}$ and $\sigma^{\ell}_{\nu,\mathrm{rel}}$ are varied from $0.01°$ to $0.1°$ and from $0.01\,\mathrm{m}$ to $0.1\,\mathrm{m}$, respectively. The color bar is scaled uniformly within each system for clear visualization.

to those in Section 6.2.1.

Fig. 8 shows the performance of the conventional and glocal fusion systems under varying N_{ws} and N_{ℓ}, respectively, along with the offline batch optimization result as a reference. In conventional fusion, performance improves with increasing N_{ws}, and a similar trend is observed in glocal fusion. However, even when $N_{\ell} = 0$, glocal fusion outperforms conventional fusion with small window sizes, and in the case of global position **p**, it even surpasses the batch optimization. With N_{ℓ} set to as few as 10, glocal fusion achieves superior global and local consistency compared to nearly all conventional fusion cases. In particular, **p**(2.5)−OM consistently outperforms conventional fusion, since the proposed method optimizes a scale factor for each glocal group, unlike conventional fusion which relies on a single global scale in monocular SLAM. Consequently, for OM, the online version of conventional fusion occasionally performs better than its offline version.

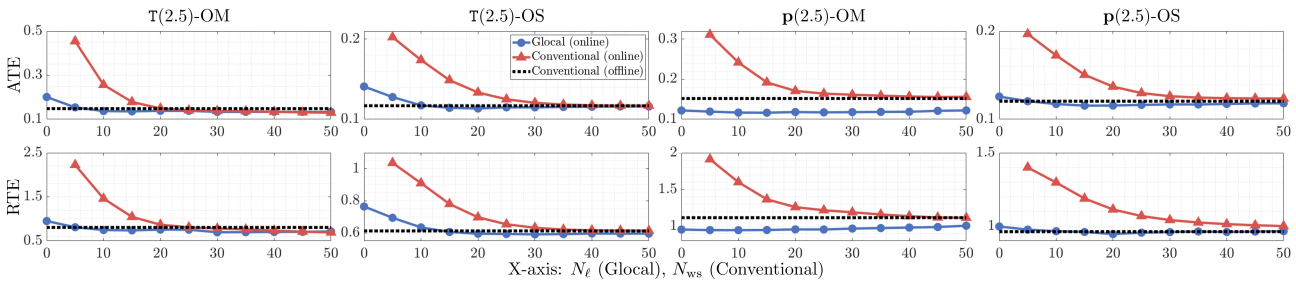

Fig. 8. Performance of conventional fusion with varying N_{ws} and glocal fusion with varying N_{ℓ}. The result of conventional fusion (offline) corresponds to batch optimization of all states without a sliding window.

Fig. 9 shows the average number of glocal groups N_{gp} generated per KITTI sequence and the average number

of local poses included in each group, as N_ℓ varies. When global pose observations are used, more groups are generated compared to the case with global position observations, since pose provides richer information and leads the glocal observer module to partition the trajectory more strictly. As N_ℓ increases, N_{gp} decreases. This indicates that larger N_ℓ enhances each group's correction capability, allowing fewer groups to preserve both global and local consistency.

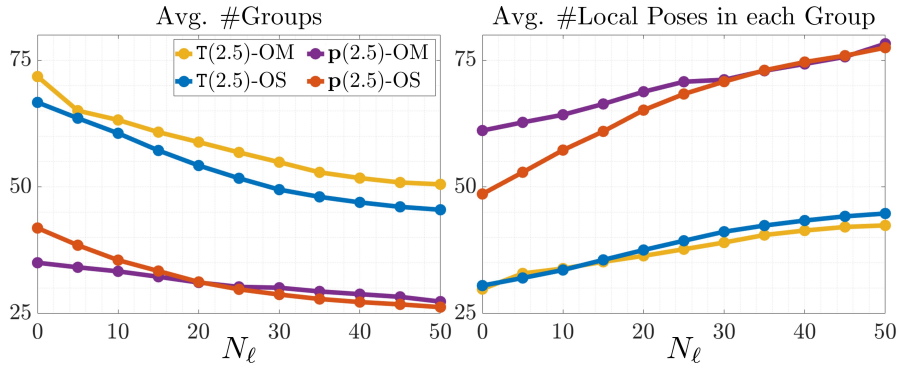

Fig. 9. Effect of the number of local states N_ℓ on group formation in the glocal fusion system. (Left) Average number of glocal groups generated per sequence. (Right) Average number of local poses assigned to each group.

Fig. 10 illustrates the results of glocal fusion (p(2.5)-OS) on KITTI seq. 9. When $N_\ell = 0$, a larger number of glocal groups are formed compared to the case of $N_\ell = 20$, resulting in noticeable discontinuities at group boundaries due to the insufficient number of local states that ensure smooth transitions. This effect is more clearly observed in the frame-wise ATE graph in Fig. 10, where group split indices cause abrupt deviations when $N_\ell = 0$. In contrast, when $N_\ell = 20$, the trajectory remains smoothly connected across groups without inter-group inconsistency, demonstrating the effective bridging role of local states designed in the glocal optimizer.

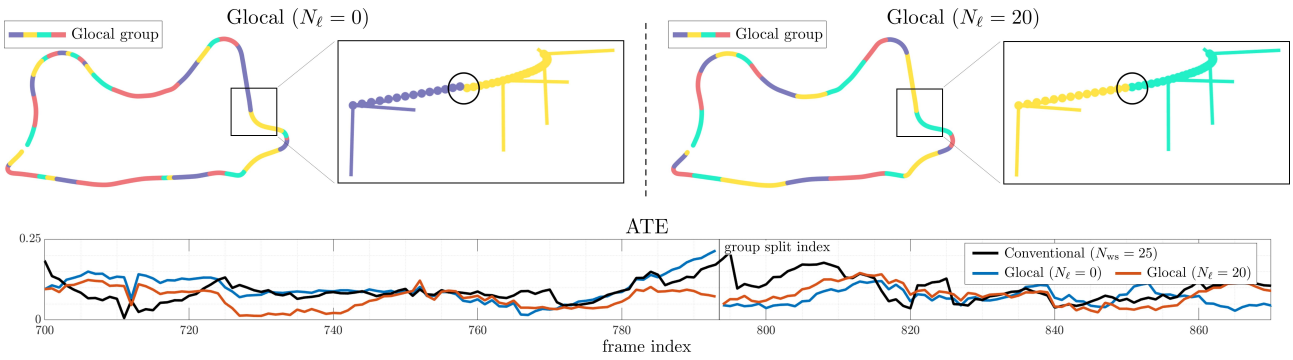

Fig. 10. Results of glocal fusion (p(2.5)-OS) on KITTI seq. 9. The top row shows estimated trajectories with $N_\ell = 0$ (left) and $N_\ell = 20$ (right). The bottom graph presents the frame-wise ATE around the highlighted group split region, comparing conventional fusion and glocal fusion with different N_ℓ values.

6.2.3. Glocal Observer

This subsection evaluates the effectiveness of the proposed glocal observer module by comparing it against a naive strategy in which glocal groups are split after a fixed number of local frames.

Fig. 11 presents the results of this comparison. In the proposed observer module, the key parameter is the local covariance $\Sigma^\ell = \text{Cov}(\sigma_\omega^\ell,\ \sigma_\nu^\ell)$, where σ_ω^ℓ and σ_ν^ℓ are varied over $\{0.01, 0.03, 0.05, 0.1, 0.15, 0.2, 0.25\}$. The resulting performance for each setting is plotted against the average number of local poses per group. As described in Section 4.2, larger values of Σ^ℓ increase the tolerance for local consistency, which allows more local frames to be included in each group, thereby shifting the x-axis values to the right.

The results show that, for a comparable number of local poses per group, the proposed observer consistently outperforms the naive fixed-frame approach. This demonstrates its ability to cluster locally consistent frames into

glocal groups, thereby achieving better accuracy with fewer parameters. The superiority becomes more evident as the group size increases, since the grouping is adaptively adjusted to preserve both global and local consistency. In addition, the module reduces redundant state variables while ensuring that only locally consistent frames are grouped, which is critical for stable optimization in large-scale trajectories.

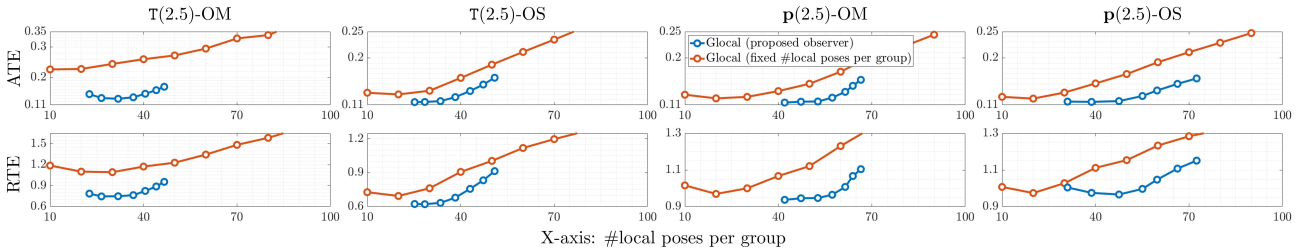

Fig. 11. Ablation study of the glocal observer module. The baseline corresponds to naive grouping with a fixed number of local frames per group. The proposed observer is evaluated by varying the local covariance Σ^{ℓ}. Results are plotted with respect to the average number of local poses per group.

6.2.4. Glocal Optimizer

The ablation study on best model selection and inter-group management in the glocal optimizer is reported in Table 2. The results indicate that incorporating the best model selection module improves performance compared to directly using the anchor node selected by the glocal observer. This suggests that the module identifies the precise indices and local states where group division should occur.

Furthermore, inter-group management achieves local and global consistency comparable to the case where all local states are included as optimization variables, while significantly reducing the number of reinstantiated states. On average, only 1.1–1.3 inter-group factors are reinstantiated, confirming that the module efficiently selects the necessary inter-group local states. Based on this observation, the average number of optimization states can be approximated as $N_{\text{ws}} + N_{\text{Reinst}} \times N_{\ell} \approx 22$.

Table 2. Ablation study of the glocal optimizer on best model selection and inter-group management (ATE/RTE in cm).

Best model selection	Inter-group management	$\mathbf{T}(2.5)$-OM			$\mathbf{T}(2.5)$-OS			$\mathbf{p}(2.5)$-OM			$\mathbf{p}(2.5)$-OS		
		ATE	RTE	N_{Reinst}	ATE	RTE	N_{Reinst}	ATE	RTE	N_{Reinst}	ATE	RTE	N_{Reinst}
✓	✓	13.03	74.80	1.19	11.74	63.29	1.16	11.67	94.67	1.30	11.74	96.59	1.31
✓		12.93	74.45	7.94	11.75	63.42	7.94	11.65	94.79	7.39	11.72	96.58	7.63
	✓	15.29	80.50	1.18	12.47	66.81	1.17	12.19	96.53	1.31	12.29	98.94	1.31

6.3. Global Observation

The fusion systems are evaluated under varying noise levels of global observations, as well as in the presence of outliers. The average performance of each system is reported over all odometry sequences of the KITTI dataset.

6.3.1. Performance Analysis with Varying Noise Level

The noise level $N_{\text{g-lvl}}$ of the global observations is varied from 0.5 to 5, and the performance of the glocal and conventional fusion systems is evaluated. The results are shown in Fig. 12. Across almost all noise levels, the glocal fusion system consistently outperforms the conventional counterpart, with the performance gap becoming more pronounced as the noise increases. This improvement can be attributed to the separation of high-noise global factors from locally consistent states, as well as the use of bundled global observations rather than single factors. When the noise level is very low, both systems achieve comparable performance. In this case, highly reliable global observations help correct the trajectory without degrading the local consistency, thereby diminishing the relative advantage of the proposed method.

Furthermore, the performance gap is larger for global position observations than for global pose observations, and the reason can be analyzed as follows. In conventional sliding-window optimization, the states included in the

window often correspond to straight-line motion. As mentioned earlier, when only global position observations are available during such motion, the observability of the rotation angle along the motion direction becomes severely limited, and the orientation must be constrained solely by the prior factor. This imbalance of information regarding the orientation between the prior factor and the current factors in the graph makes the conventional fusion system highly sensitive to small amounts of noise. In contrast, the glocal fusion system can optimize a much longer trajectory history within the same window size, thereby alleviating this issue. Consequently, when global position observations are used, the performance of the conventional approach degrades compared to the proposed method.

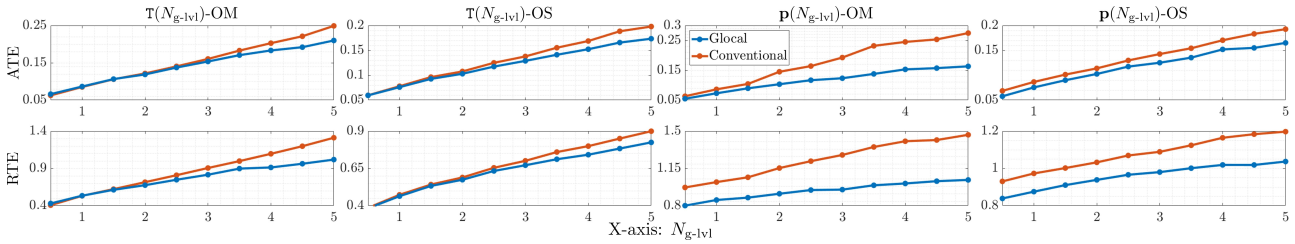

Fig. 12. Comparison of glocal and conventional fusion systems under varying noise levels of global observations.

Fig. 13 illustrates the variation in the number of glocal groups and the average number of local poses per group as functions of the global noise level. As the noise level decreases, the glocal observer can more strictly detect glocal errors using higher-quality global observations, which leads to the formation of a larger number of groups.

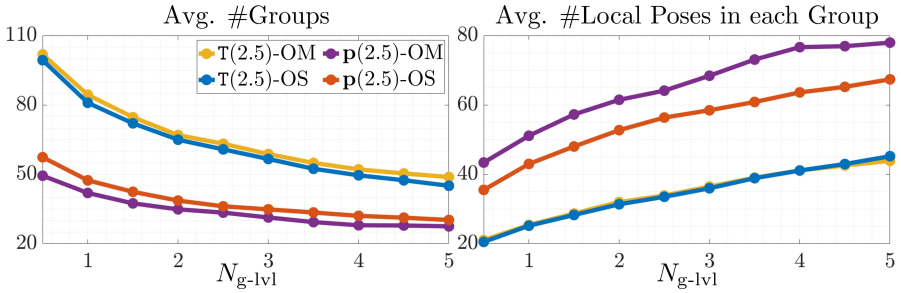

Fig. 13. Effect of the global noise level $N_{g\text{-}lvl}$ on group formation.

6.3.2. Outlier

This subsection evaluates the performance of each fusion system in the presence of outliers in the global observations. The covariance corresponding to a noise level of 2.5 is used for generating inliers, while outliers are generated with a new covariance corresponding to a noise level of 5. For the outlier data, the covariance corresponding to a noise level of 2.5 is assigned, thereby introducing a mismatch between the observation and its covariance model. The outlier ratio among the global observations is varied from 10% to 50%, and the outliers are uniformly distributed across the dataset.

The results obtained with global observations containing outliers are presented in Fig. 14. As the outlier ratio increases, the proposed glocal fusion system demonstrates superior performance compared to the conventional fusion system. This indicates that the parameterization adopted in glocal fusion propagates covariance information more effectively than the conventional approach, which relies on individual local states, thereby enabling more robust outlier detection and rejection.

6.3.3. Degeneracy Case without Global Observations

In this subsection, the performance of the fusion systems is evaluated in scenarios where no global observations are available for certain segments of the sequence. The evaluated fusion cases are $\mathbf{p}(2.5)$-OM and $\mathbf{p}(2.5)$-OS, tested on KITTI seq. 7. The results are illustrated in Fig. 15. In the trajectory, areas A and B represent segments where global observations are absent until new observations are received. In both regions, the trajectory estimated by the glocal fusion system remains close to the true trajectory, whereas the trajectory from the conventional fusion system shows

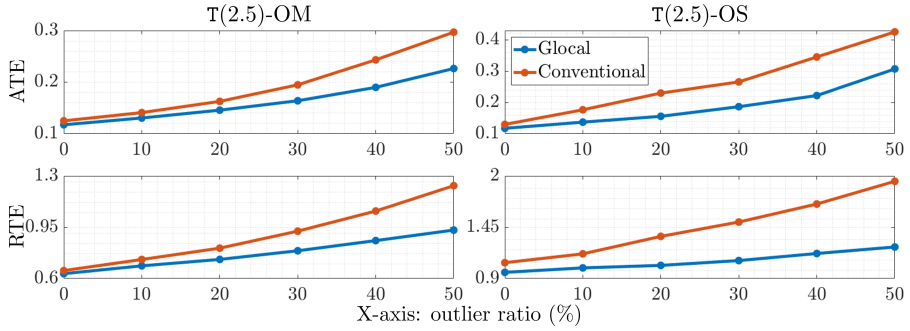

Fig. 14. Results under different outlier ratios in the global observations.

a significant loss of global consistency. This discrepancy is particularly pronounced in **p**(2.5)−OM with monocular SLAM, where the absence of scale optimization results in a large deviation. A similar trend is observed in **p**(2.5)−OS, where the glocal fusion again achieves more accurate results.

The ATE graphs at the bottom of Fig. 15 highlight the difference more clearly: while the conventional fusion system completely loses global consistency in the no-global-observation regions, the glocal fusion system maintains a high level of consistency. This is attributed to the fact that the glocal fusion approach continuously optimizes an extended trajectory history, allowing global observations from neighboring groups to propagate and restore global consistency even in segments without direct global observations.

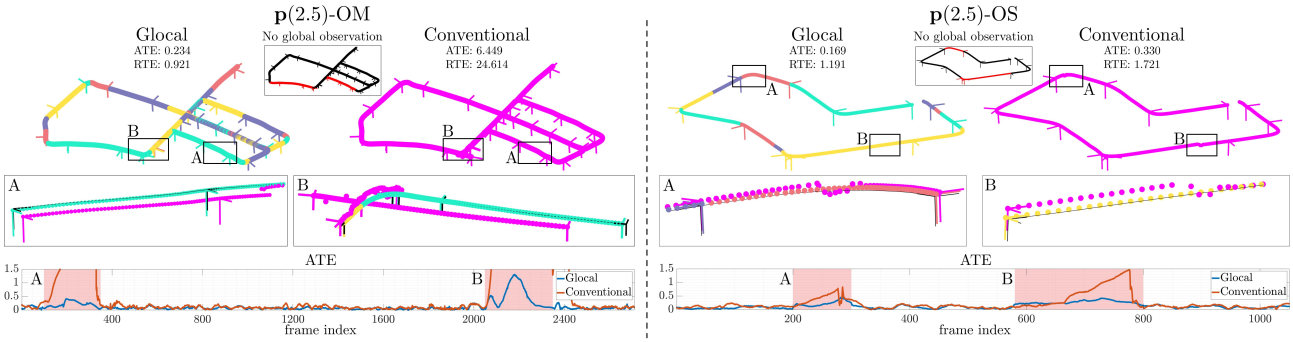

Fig. 15. Results of **p**(2.5)−OM and **p**(2.5)−OS fusion cases on KITTI seq. 7. In both cases, the regions marked as A and B correspond to trajectory segments without global observations. In the zoomed-in views of regions A and B, the ground truth is represented by the black trajectories.

6.4. Experimental Results

6.4.1. KITTI

The evaluation results for all KITTI sequences are presented in Fig. 16. Each fusion system is tested using global pose and position observations with a noise level of 2.5, combined with four types of local SLAM. In almost all cases, the proposed glocal fusion system achieves superior preservation of both global and local consistency compared to the conventional approach. The averages of ATE and RTE across each fusion case also confirm the improved performance of the proposed system. Notably, the performance difference is more pronounced in local consistency than in global consistency. Fig. 17 shows the qualitative results of **p**(2.5)−LM together with the number of glocal groups formed in each sequence.

Fusion	T(2.5)-OM				T(2.5)-OS				T(2.5)-VF				T(2.5)-LM			
method	Glocal		Conven.		Glocal		Conven.		Glocal		Conven.		Glocal		Conven.	
seq.	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE
0	13.76	79.38	13.79	81.13	12.05	72.52	12.49	74.68	11.24	66.80	12.04	69.57	10.59	61.69	11.49	66.11
1	-	-	-	-	16.75	64.22	17.96	64.24	52.18	224.61	54.38	238.92	12.17	50.08	12.78	44.74
2	13.25	73.21	13.92	75.48	13.55	83.76	14.41	83.77	12.68	69.52	12.75	68.51	38.38	163.65	67.64	253.37
3	10.43	63.46	10.85	65.02	10.27	61.76	10.64	63.30	10.53	62.43	10.89	64.78	10.12	59.62	10.33	60.81
4	14.40	72.65	13.91	61.95	12.35	53.99	12.22	52.44	15.57	69.82	16.21	69.73	11.88	50.21	12.05	51.14
5	11.26	70.64	12.58	78.79	9.58	56.51	10.85	61.49	9.77	56.42	11.34	63.99	8.64	47.89	10.51	57.89
6	14.89	84.03	20.60	124.31	10.22	54.92	11.52	59.02	11.13	59.70	12.53	65.26	8.80	43.97	10.93	54.29
7	12.56	79.37	13.92	96.21	9.69	58.65	10.39	62.14	9.37	58.56	10.59	64.56	8.67	51.35	10.03	59.60
8	14.35	86.64	14.96	87.58	12.55	72.54	12.87	73.79	12.69	71.87	13.22	74.66	12.58	69.90	12.84	72.01
9	13.47	67.21	13.27	67.35	11.77	59.87	12.29	60.73	13.43	66.82	13.96	70.09	10.42	53.21	11.41	56.50
10	11.90	71.39	12.64	70.85	10.33	57.48	11.80	63.81	10.69	62.16	11.84	66.02	10.01	55.77	11.59	62.89
Avg.	13.03	74.80	14.05	80.87	11.74	63.29	12.50	65.40	15.39	78.97	16.34	83.28	12.93	64.30	16.51	76.31

Fusion	p(2.5)-OM				p(2.5)-OS				p(2.5)-VF				p(2.5)-LM			
method	Glocal		Conven.		Glocal		Conven.		Glocal		Conven.		Glocal		Conven.	
seq.	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE	ATE	RTE
0	10.72	111.24	13.30	130.19	11.55	114.97	11.99	121.32	10.86	133.53	11.70	151.19	10.34	104.99	11.27	112.44
1	-	-	-	-	14.02	135.20	16.32	144.25	56.19	260.10	57.08	263.86	10.24	102.03	11.82	108.28
2	13.06	114.68	13.33	124.12	14.39	120.58	14.83	137.33	13.26	116.40	13.23	140.94	52.73	506.54	78.58	612.53
3	10.23	91.84	11.37	99.24	9.95	90.76	10.81	98.13	11.13	100.56	11.58	105.48	9.93	121.23	10.69	126.97
4	16.62	88.34	42.21	124.95	14.10	76.02	17.67	77.28	31.15	116.78	132.84	168.40	15.98	73.79	35.49	68.67
5	9.22	75.66	12.71	110.25	9.84	76.72	11.12	92.02	10.94	120.93	11.50	151.11	8.33	71.13	10.62	86.39
6	10.67	75.18	16.68	130.05	10.21	75.72	11.76	83.02	12.32	84.83	12.95	94.79	9.22	73.25	11.40	81.97
7	10.48	84.92	13.91	118.56	9.74	81.97	11.11	95.72	10.53	138.58	11.55	163.86	9.23	73.03	10.92	87.64
8	12.98	122.97	14.51	148.06	12.55	116.73	12.89	125.63	12.99	117.23	13.33	140.40	12.41	112.73	12.70	126.92
9	11.74	86.82	13.66	120.95	11.76	84.64	12.54	97.41	13.28	99.20	13.73	118.76	10.69	85.14	11.68	97.41
10	10.93	95.07	12.33	111.59	11.03	89.20	11.72	104.13	11.07	116.58	11.90	144.29	10.77	79.39	11.51	91.18
Avg.	11.67	94.67	16.40	121.80	11.74	96.59	12.98	106.93	17.61	127.70	27.40	149.37	14.53	127.57	19.70	145.49

Fig. 16. Fusion results on all KITTI sequences (ATE/RTE in cm).

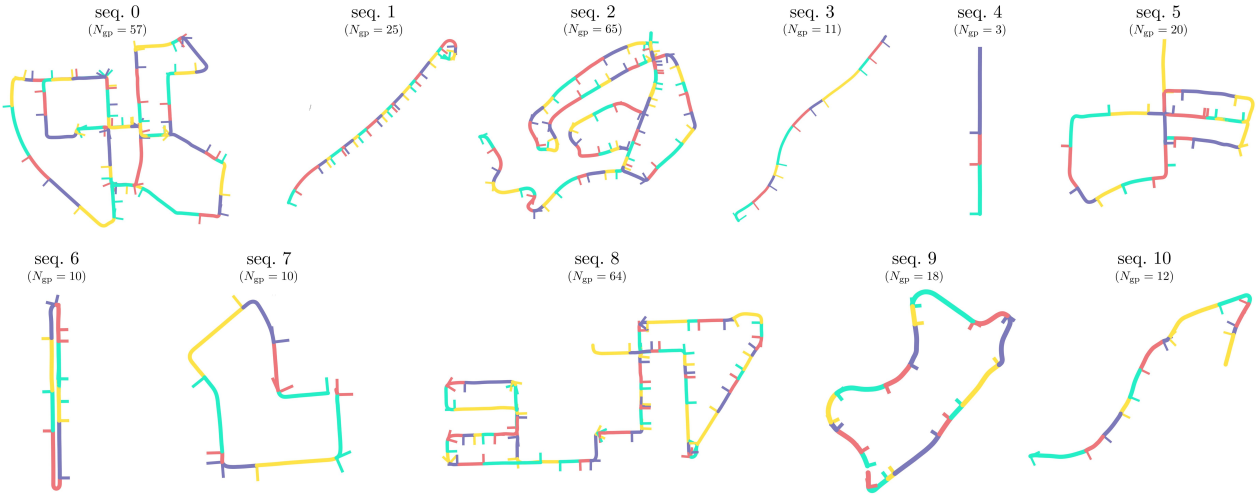

Fig. 17. Qualitative results of glocal fusion (p(2.5)-LM), along with the number of glocal groups formed in each sequence.

6.4.2. Runtime Analysis

In this subsection, the runtime of the key modules in the glocal and conventional fusion systems is analyzed. The results are summarized in Table 3. The statistics are obtained from all sequences, executed in MATLAB on a desktop (CPU: Intel i9-10900K @ 3.7GHz, RAM: 32GB), and normalized with respect to the time of the conventional fusion system for comparison. The glocal observer runs 30–40 times faster than the conventional optimization, demonstrating its effectiveness in detecting glocal errors. Although the best-model selection and optimization in the glocal optimizer require approximately 1–2 times longer than conventional optimization, the optimizer is triggered much less frequently: on average every 32.41 frames with global pose observations, and about every 60 frames with global position observations.

As a result, the overall system achieves optimization of the entire trajectory in an online manner, while the total runtime is approximately 7–10 times faster than the conventional approach. This efficiency arises because the observer triggers group formation and optimization only when necessary, while otherwise performing glocal error checks and updating the current glocal state at high speed, leading to both faster and more effective fusion than the conventional system.

Table 3. Runtime analysis of the key modules for the glocal and conventional fusion systems (median $\pm 2\times$ std. deviation).

	Glocal Fusion	
Global obs. type	T	p
Glocal observer	0.031 ± 0.126	0.027 ± 0.171
Glocal optimizer	(trig. / 32.41 fr.)	(trig. / 59.28 fr.)
ㄴ best model sel.	1.950 ± 1.328	1.950 ± 1.328
ㄴ optimization	1.061 ± 1.244	1.405 ± 3.757
Total runtime	0.135	0.094

	Conventional Fusion	
Global obs. type	T	p
Global alignment	0.066 ± 0.063	0.034 ± 0.047
optimization	1.000 ± 0.343	1.000 ± 0.653
Total runtime	1.000	1.000

*Module runtimes are normalized by the conventional fusion's optimization runtime.
*The total runtimes are normalized by that of the conventional fusion system.

7. Conclusion

This paper presented a glocal fusion framework for robot navigation that simultaneously ensures global and local consistency. The proposed approach analyzed the characteristics of local SLAM systems and introduced a novel parameterization of glocal groups and states, enabling a new factor graph structure for sensor fusion. This design prevents noisy global observations from contaminating strongly consistent local states and provides robustness to covariance modeling through the natural separation of heterogeneous local and global factors. Furthermore, by significantly reducing the number of optimization variables, the glocal fusion approach achieves a structure well suited for online navigation. Based on this approach, a glocal navigation system is developed, consisting of two core modules: a glocal observer, which rapidly detects glocal errors, and a glocal optimizer, which effectively restores both global and local consistency.

The advantages of the proposed framework are validated through extensive experiments. Using four representative local SLAM systems and global observations with varying noise levels, outliers, and missing data, the glocal fusion system is compared against conventional fusion systems. The results demonstrated that the proposed system consistently preserves higher levels of both global and local consistency, and shows markedly superior performance to conventional methods under conditions of high noise, outlier contamination, or interruptions of global observations. In addition, by reducing the number of optimization parameters and performing glocal optimization only when necessary, the proposed system achieved highly faster performance compared to conventional systems.

The proposed glocal fusion framework can be applied to a wide range of systems that integrate global and local observations. Future work will focus on extending this framework to tightly coupled designs that incorporate diverse sensor modalities for enhanced robustness and performance.

Reference

[1] R. Mur-Artal and J. D. Tardos, “Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras,” IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255–1262, 2017.

[2] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardos, “Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam,” IEEE Transactions on Robotics, vol. 37, no. 6, pp. 1874–1890, 2021.

[3] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 3, pp. 611–625, 2017.

[4] J. Zhang, S. Singh et al., “Loam: Lidar odometry and mapping in real-time,” in Robotics: Science and Systems, vol. 2, no. 9, Berkeley, CA, 2014, pp. 1–9.

[5] T. Shan and B. Englot, “Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain,” in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018, pp. 4758–4765.

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

[7] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, “A robust and modular multi-sensor fusion approach applied to mav navigation,” in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2013, pp. 3923–3929.

[8] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, “Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft mav,” in 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2014, pp. 4974–4981.

[9] J. E. Yoder, P. A. Iannucci, L. Narula, and T. E. Humphreys, “Multi-antenna vision-and-inertial-aided cdgnss for micro aerial vehicle pose estimation,” in Proceedings of the ION GNSS+ Meeting, Online, 2020, pp. 2281–2298.

[10] T. Qin, S. Cao, J. Pan, and S. Shen, “A general optimization-based framework for global pose estimation with multiple sensors,” arXiv preprint arXiv:1901.03642, 2019.

[11] R. Mascaro, L. Teixeira, T. Hinzmann, R. Siegwart, and M. Chli, “Gomsf: Graph-optimization based multi-sensor fusion for robust uav pose estimation,” in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 1421–1428.

[12] X. Niu, H. Tang, T. Zhang, J. Fan, and J. Liu, “Ic-gvins: A robust, real-time, ins-centric gnss-visual-inertial navigation system,” IEEE Robotics and Automation Letters, vol. 8, no. 1, pp. 216–223, 2022.

[13] S. Cao, X. Lu, and S. Shen, “Gvins: Tightly coupled gnss–visual–inertial fusion for smooth and consistent state estimation,” IEEE Transactions on Robotics, vol. 38, no. 4, pp. 2004–2021, 2022.

[14] C. Wang, H. Zhang, T.-M. Nguyen, and L. Xie, “Ultra-wideband aided fast localization and mapping system,” in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017, pp. 1602–1609.

[15] A. Arun, R. Ayyalasomayajula, W. Hunter, and D. Bharadia, “P2slam: Bearing based wifi slam for indoor robots,” IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 3326–3333, 2022.

[16] K. Ismail, R. Liu, Z. Qin, A. Athukorala, B. P. L. Lau, M. Shalihan, C. Yuen, and U.-X. Tan, “Efficient wifi lidar slam for autonomous robots in large environments,” in 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE). IEEE, 2022, pp. 1132–

1137.
[17] J. Zhang, Y. Sun, J. Huai, Y. Wang, C. Wang, and L. Xie, “Vins-uwb: Robust and accurate navigation aided by uwb measurements,” IEEE Transactions on Automation Science and Engineering, vol. 19, no. 4, pp. 3080–3095, 2021.
[18] M. Yang, C. Yuan, Q. Zhang, T. Wu, and C. Wu, “Ultra tightly coupled gnss/uwb/ins integration for resilient autonomous vehicle localization in urban areas,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–16, 2021.
[19] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” Proceedings of the IEEE, vol. 92, no. 3, pp. 401–422, 2004.
[20] K. Konolige, G. Grisetti, R. Kummerle, W. Burgard, B. Limketkai, and R. Vincent, “Efficient sparse pose adjustment for 2d mapping,” in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2010, pp. 22–29.
[21] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, “A tutorial on graph-based slam,” IEEE Intelligent Transportation Systems Magazine, vol. 2, no. 4, pp. 31–43, 2010.
[22] K. M. Wurm, C. Stachniss, and W. Burgard, “Coordinated multi-robot exploration using a segmentation of the environment,” in 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2008, pp. 1160–1165.
[23] J. Knuth and R. Sukhatme, “Improving multi-robot coordination by learning the global utility map,” in 2006 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2006, pp. 3458–3463.
[24] J. L. Blanco, J.-A. Fernández-Madrigal, and J. González, “A novel measure of uncertainty for mobile robot slam with rao-blackwellized particle filters,” The International Journal of Robotics Research, vol. 27, no. 1, pp. 73–89, 2008.
[25] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. John Wiley & Sons, 2004.
[26] J. A. Castellanos, J. Neira, and J. D. Tardos, “Limits to the consistency of ekf-based slam,” in Proceedings 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 1. IEEE, 2004, pp. 763–769.
[27] R. I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.
[28] T. D. Barfoot, State Estimation for Robotics. Cambridge University Press, 2017.
[29] S. Huang and G. Dissanayake, “Convergence and consistency analysis for extended kalman filter based slam,” IEEE Transactions on Robotics, vol. 23, no. 5, pp. 1036–1049, 2007.
[30] S. Thrun and Y. Liu, “Multi-robot slam with sparse extended information filers,” in Robotics Research. Springer, 2005, pp. 254–266.
[31] M. Kaess, A. Ranganathan, and F. Dellaert, “iSAM: Incremental smoothing and mapping,” IEEE Transactions on Robotics, vol. 24, no. 6, pp. 1365–1378, 2008.
[32] F. Dellaert, “Factor graphs and GTSAM: A hands-on introduction,” Georgia Institute of Technology, Tech. Rep., 2012.
[33] J. Dong, X. Zhou, and R. Tedrake, “Sparse incremental gaussian mixture models,” in 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017, pp. 6337–6344.
[34] Y. Tian, H. Wang, R. Guo, and S. Shen, “Vdo-slam: A visual dynamic object-aware slam system,” in 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020, pp. 6093–6099.

[35] H. Liu, M. Long, J. Wang, and M. Jordan, “Transferable representation learning with deep adaptation networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 12, pp. 3071–3085, 2018.
[36] S. Agarwal, K. Mierle, and Others, “Ceres solver,” http://ceres-solver.org, 2012.
[37] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
[38] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in European Conference on Computer Vision (ECCV). Springer, 2006, pp. 404–417.
[39] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in 2011 International Conference on Computer Vision (ICCV). IEEE, 2011, pp. 2564–2571.
[40] J. Shi and C. Tomasi, “Good features to track,” in 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 1994, pp. 593–600.
[41] B. D. Lucas, T. Kanade et al., “An iterative image registration technique with an application to stereo vision,” in IJCAI, vol. 81, 1981, pp. 674–679.
[42] B. K. Horn, “Closed-form solution of absolute orientation using unit quaternions,” JOSA A, vol. 4, no. 4, pp. 629–642, 1987.