On Bundle Adjustment for Multiview Point Cloud Registration

Huaiyang Huang, Yuxiang Sun, Jin Wu, Jianhao Jiao, Xiangcheng Hu, Linwei Zheng, Lujia Wang, and Ming Liu

***Abstract*—Multiview registration is used to estimate Rigid Body Transformations (RBTs) from multiple frames and reconstruct a scene with corresponding scans. Despite the success of pairwise registration and pose synchronization, the concept of Bundle Adjustment (BA) has been proven to better maintain global consistency. So in this work, we make the multiview point-cloud registration more tractable from a different perspective in resolving range-based BA. We first analyse the optimal condition of the objective function of BA that unifies some previous approaches. Based on this analysis, we propose an objective function that takes both measurement noises and computational cost into account. For the feature parameter update, instead of calculating the global distribution parameters from the raw measurements, we aggregate the local distributions in a frame-wise fashion at each iteration. The computational cost of feature update is then only dependent on the number of scans. Finally, we develop a multiview registration system using voxel-based quantization that can be applied in real-world scenarios. The experimental results demonstrate our superiority over the baselines in terms of both accuracy and speed. Moreover, the results also show that our average positioning errors achieve the centimeter level. Related materials are available at our project page https://hyhuang1995.github.io/bareg/.**

***Index Terms*—SLAM, state estimation, robot sensing systems, bundle adjustment, point cloud registration.**

I. Introduction

POINT cloud registration has long been an essential research problem in the computer vision and robotics community [1]. Particularly for autonomous navigation, the availability of range sensors capable of efficiently reconstructing the 3D structure has significantly improved the systematic performance [2]. This fact stimulates the research on multiview registration, which combines multiple scans of point-cloud data (PCD) into a globally consistent structural model [3].

In the early stage, the multiview registration problem is resolved by either performing pairwise registration [4], [5], or combining this scheme with pose synchronization [6]. For decades, pairwise registration is the most prevailing paradigm and is widely applied in different state estimation systems for range sensors [5], [7]. To tackle this problem in the multiview regime, the systems are generally designed in a frame-to-frame or frame-to-model fashion. The problem is thus simplified into resolving the Rigid Body Transformation (RBT) concerning the latest scan. Despite the prevalence and success of various pairwise registration methods, global consistency is less considered in these implementations.

On the top of pairwise registration methods, some research work explicitly deals with the global consistency with the introduction of pose synchronization [8], [9]. This is achieved by first registering pairs of overlapped scans and then optimizing a pose graph from the registered relative poses. To this end, recent research resorts to the Bundle Adjustment (BA), a concept that originated from visual Structure-from-Motion (SfM) that simultaneously estimates the frame poses and the position of visual landmarks. Nevertheless, compared to visual state estimation applications, relatively fewer discussions on leveraging inter-frame constraints can be found for systems based on range sensors. Some methods are proposed to fully exploit the inter-frame constraints provided by a sequence of point-cloud data. Firstly, Landmark-based methods parameterize the local measurements as individual landmarks. Similar to visual BA, landmark and pose parameters are jointly optimized [10], [11]. These methods generally require specific feature detection, and the measurement noises are not easy to be taken into account. Most recently, two works propose to directly use the eigenvalue of sample covariance as the objective function [12], [13], which we named as EigenValue Minimization (EVM)-based formulation. These methods provide an elegant formulation for multiview registration, while the trade-off still exists between the generality in formulation and performance in applications.

In this paper, we first analyze the optimal condition of EVM, which unifies the Landmark and the EVM-based method. Then, based on the above analysis, we derive an objective function for resolving multiview registration, which can be applied to a standard least-square problem for more efficient optimization. Different from previous methods, this formulation also fully

Manuscript received March 25, 2021; accepted July 29, 2021. Date of publication August 18, 2021; date of current version September 1, 2021. This letter was recommended for publication by Associate Editor N. Kottege and Editor P. Pounds upon evaluation of the reviewers' comments. This work was supported in part by Zhongshan Municipal Science, and Technology Bureau Fund, under Project ZSST21EG06, in part by Collaborative Research Fund by the Research Grants Council Hong Kong, under Project C4063-18G, and in part by the Department of Science and Technology of Guangdong Province Fund, under Project GDST20EG54, awarded to Prof. Ming Liu. *(Corresponding author: Ming Liu.)*

Huaiyang Huang, Jin Wu, Jianhao Jiao, Xiangcheng Hu, Linwei Zheng, Lujia Wang, and Ming Liu are with the Robotics Institute, The Hong Kong University of Science and Technology, Hong Kong, China (e-mail: hhuangat@connect.ust.hk; jin_wu_uestc@hotmail.com; jiaojh1994@gmail.com; 2022087641@qq.com; lzhengad@connect.ust.hk; rugga.wang@gmail.com; liu.ming.prc@gmail.com).

Yuxiang Sun is with the Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: yx.sun@polyu.edu.hk, sun.yuxiang@outlook.com).

Digital Object Identifier 10.1109/LRA.2021.3105686

considers the measurement noise. Finally, we provide a systematic implementation that can be applied in real-world scenarios. We summarize our contributions as follows:

1) We provide an analysis on the optimal condition of EVM-formulation for resolving multiview registration, which unifies the methods by extracting landmarks and directly optimizing eigenvalues.
2) We propose a novel objective function based on the analysis, which takes account of both computational efficiency and the measurement noise.
3) We develop a voxel-based multiview registration pipeline with the proposed objective function and local distribution aggregation.

II. Related Works

A. Pairwise Registration

Pairwise registration is the most prevailing paradigm in the range-based state estimation. This track of methods considers the only variable to optimize is the pose of the latest frame. To estimate, for example, the states of a sequence of point-cloud data, they generally can be divided into two categories, i.e., frame-to-frame and frame-to-model methods.

Frame-to-frame registration has long been an important problem in robotic perception. Among these works, ICP [7] is one of the most popular pipelines, along with its variants [14], [15]. In an Expectation-Maximum (EM) scheme, they find the correspondences with the pose estimation and optimize the pose parameters with the association result. As an extension to frame-to-frame method, the frame-to-model pipeline is widely applied in SLAM [5], [16]–[18] and localization [19], [20] systems for robotic state estimation. For example, in LOAM [5], scans in the past frames are aggregated into a global map with their estimated poses. After every frame-to-frame registration step, the system refines the pose estimation with the downsampled global map. Some work [5], [21] shows that the frame-to-model refinement significantly improves the localization and mapping performance. Despite the success and brevity of these methods, the concept of BA is seldom exploited, unlike vision-based pipelines.

B. Multiview Registration

In contrast to pairwise registration, approaches targeting multiview registration take multiple frames of PCD that are partially overlapped as input. With local observations cross different frames, they simultaneously optimize the poses of different frames in multiple point-cloud frames.

In the early stage, the majority of methods are based on pairwise registration and pose synchronization [3], [8], [22]–[24]. This formulation is not directly derived from the sensor measurement model, thus it is more suitable for producing initial estimation for further refinement. Another track of methods parameterizes the measurements as different geometric landmarks [10], [11], [25], namely the Landmark-base methods. Similar to previous methods, the formulation does not consider the measurement noise, and generally extracting geometric features are required in advance.

Most recently, some work provides a different perspective to this problem [12], [13]. They directly optimize the eigenvalues of the local distribution parameters for a global consistent solution. In this paper, we further discuss the optimal condition of this formulation. In addition, different from previous works, we show that the covariance can be aggregated in frame-wise and provide a objective function that can be applied to a least-squares solver.

III. Background

A. Notations

We use bold uppercase $\mathbf{H}$ for the matrices, bold lowercase $\mathbf{x}$ for the vector, and light lower case (e.g. θ) for the scalar. For a matrix $\mathbf{H}$, its eigenvalues are denoted by $\lambda_i(\mathbf{H})$, which are enumerated in a descending order: $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. For $\mathbf{H} \in \mathbb{R}^{3\times 3}$ and $\mathbf{H}$ is a symmetric matrix, λ_i can be calculated from Singular Value Decomposition (SVD) by $\mathbf{H} = \mathbf{R_H}\mathbf{\Lambda_H}\mathbf{R_H}^T$, where we have

$$\mathbf{\Lambda_H} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3), \mathbf{R_H} \in \mathrm{SO}(3). \tag{1}$$

The pose of k-th frame $\mathcal{F}_k$ is represented by the RBT $\mathbf{T}_k = (\mathbf{R}_k, \mathbf{t}_k)$, where $\mathbf{R}_k \in \mathrm{SO}(3)$ and $\mathbf{t}_k \in \mathbb{R}^3$. A 3D point $\mathbf{p}_i$ expressed under $\mathcal{F}_k$ can be transformed to the global frame $\mathcal{W}$ using: $\mathbf{R}_k\mathbf{p}_i + \mathbf{t}_k$. $\mathbf{T}_k$ can be further parameterized as a vector via $\boldsymbol{\xi}_k = \log(\mathbf{T}_k{}^\vee), \boldsymbol{\xi}_k \in \mathfrak{se}(3)$ and expressed back via $\mathbf{T}_k = \exp(\boldsymbol{\xi}_k{}^\wedge)$.

B. Problem Formulation

Suppose a common feature is observed by a set of frames $\{\mathcal{F}_1, \ldots, \mathcal{F}_N\}$ and the feature is parameterized by $\boldsymbol{\pi}$ under $\mathcal{W}$. Then the total state vector can be defined by $\mathbf{x} = [\boldsymbol{\xi}_1, \ldots \boldsymbol{\xi}_N, \boldsymbol{\pi}]$. The local observation of this feature in $\mathcal{F}_k$ is represented by a set of points $\mathcal{P}_k = \{\mathbf{p}_{k1}, \ldots, \mathbf{p}_{kn_k}\}$, and the geometric constraint is modeled by the measurement function: $r(\boldsymbol{\xi}_k, \mathbf{p}_{ki}, \boldsymbol{\pi})$. Taking the planar feature as an example, if we parameterize the planar feature as $\boldsymbol{\pi} = [\mathbf{n}, \boldsymbol{\mu}]$, where $\mathbf{n}$ is the normal of the plane and $\boldsymbol{\mu}$ can be an arbitrary point on the surface. For the simplicity of the following discussions, we define $\boldsymbol{\mu}$ as the sample mean. Accordingly, the geometric constraint is the point-to-plane distance:

$$r\left(\boldsymbol{\xi}_k, \mathbf{p}_{ki}, \boldsymbol{\pi}\right) = \mathbf{n}^T\left(\mathbf{R}_k\mathbf{p}_{ki} + \mathbf{t}_k - \boldsymbol{\mu}\right). \tag{2}$$

With the Gaussian noise assumption, the objective of multiview registration problem is to minimize the following energy function:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{k=1}^{N} \frac{1}{n_k} \sum_{i=1}^{n_k} \left\| r\left(\boldsymbol{\xi}_k, \mathbf{p}_{ki}, \boldsymbol{\pi}\right)\right\|_2^2. \tag{3}$$

Methods for solving this optimization problem are described in the following section.

C. Resolving Multiview Registration

Here we provide a brief introduction to the previous methods as for the background of further discussions.

Planar Landmark (PL) This track of methods estimates feature parameters from the local measurements and then simultaneously optimizes the feature and pose parameters. This scheme is very similar to the pipeline of visual BA. There are diverse methods for feature parameterization and estimation. For example, Principle Component Analysis (PCA) is a widely used technique for planar feature extraction. Given the set of local measurements $\{\mathbf{p}_{k1}, \ldots, \mathbf{p}_{kn_k}\}$ under $\mathcal{F}_k$, sample mean and covariance is given as:

$$\begin{cases} \boldsymbol{\mu}_k^\ell = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{p}_{ki} \\ \boldsymbol{\Sigma}_k^\ell = \frac{1}{n_k} \sum_{i=1}^{n_k} \left(\mathbf{p}_{ki} - \boldsymbol{\mu}_k^\ell\right) \left(\mathbf{p}_{ki} - \boldsymbol{\mu}_k^\ell\right)^T \end{cases}, \tag{4}$$

And the surface normal is estimated from Singular Value Decomposition (SVD) of $\boldsymbol{\Sigma}_k^\ell$. To leverage inter-frame constraints, generally, a distance function on $\boldsymbol{\pi}$ is defined, which is appended to the objective function (3) in the optimization.

BALM Given the current estimated poses, the points under global coordinate can be aggregated as $\mathcal{P} = \{\mathbf{R}_{ki}\mathbf{p}_{ki} + \mathbf{t}_k | k = 1, \ldots, N, i = 1, \ldots n_k\}$. The sample mean and covariance of $\mathcal{P}$ are denoted by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, given by:

$$\begin{cases} \boldsymbol{\mu} = \frac{1}{\sum_k n_k} \sum_{i=1}^{n_k} \mathbf{p}'_{ki} \\ \boldsymbol{\Sigma} = \frac{1}{\sum_k n_k} \sum_{i=1}^{n_k} \left(\mathbf{p}'_{ki} - \boldsymbol{\mu}\right) \left(\mathbf{p}'_{ki} - \boldsymbol{\mu}\right)^T \end{cases}, \tag{5}$$

where $\mathbf{p}'_{ki} = \mathbf{R}_k\mathbf{p}_{ki} + \mathbf{t}_k$.

Lemma: 1 Assume known optimal feature parameters $\boldsymbol{\pi}$, the objective function is equivalent to minimize the minimal eigenvalue, that is:

$$\lambda_3\left(\boldsymbol{\Sigma}\right) = \frac{1}{\sum_k n_k} \sum_{i=1}^{n_k} \left\| \mathbf{n}^T \left(\mathbf{R}_k\mathbf{p}_{ki} + \mathbf{t}_k - \boldsymbol{\mu}\right) \right\|_2^2. \tag{6}$$

Proof: We refer the readers to BALM [13] and [26].

Based on *Lemma. 1*, BALM directly resolves the multiview registration problem via:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \lambda_3\left(\boldsymbol{\Sigma}\right). \tag{7}$$

We name this formulation as EigenValue Minimization (EVM) formulation. Assuming the optimal feature parameter $\boldsymbol{\pi}$ is calculated in advance of the optimization, this formulation is only dependent on the frame poses.

Eigen-Factor (EF) EF uses Homogenous point representation $\tilde{\mathbf{p}} = [\mathbf{p}, 1]$, and the local feature is parameterized as $\boldsymbol{\eta} = [\mathbf{n}, -\mathbf{n}^T\boldsymbol{\mu}]$. The objective function is re-derived into:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{k=1}^{N} \boldsymbol{\eta}^T \mathbf{T}_k \underbrace{\tilde{\mathbf{P}}_k \tilde{\mathbf{P}}_k^T}_{\mathbf{S}_k} \mathbf{T}_k^T \boldsymbol{\eta}, \tag{8}$$

where each column of $\tilde{\mathbf{P}}_k \in \mathbb{R}^{4\times n_k}$ corresponds to stacked transformed homogeneous points. Then with the first-order gradient descending method, EF optimizes this objective function to resolve the frame poses. Note that as this formulation gives the same point-to-plane distance, it is equivalent to the EVM as in (6).

Remark 1: The first issue is about the feature parameter estimation in PL-based methods. Introducing plane parameters into the optimization variables lead to a large optimization structure. This also causes information loss as the raw measurements are neglected.

Remark 2: The first issue is the handling of the optimal feature parameters. We observe that the optimal feature parameters $\hat{\boldsymbol{\pi}}$ varies with the update of frame poses $\{\mathbf{T}_k\}$. As a consequence, although the EVM formulation is independent of feature parameters, the update of feature parameters would require recomputation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Therefore, BALM assumes the optimal feature parameter is resolved before optimization, and EF uses Homogeneous representation to avoid point-wise update of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. However, as the formulation of EF is applicable only to planar features, it restricts the discussion on other geometric feature types.

Remark 3: Another trade-off exists between the formulation and the efficiency. As the original objective function is re-formulated as the minimal eigenvalue, it is no longer in a least-squares formulation. Solving it with first-order gradient descent methods is difficult to converge efficiently. On the contrary, BALM uses a second-order approximation which requires the sophisticated computation of large-scale Hessian matrices, the time complexity of which is dependent on the number of points in the input scan.

IV. Methodology

In this section, we provide theoretical analysis for the optimal condition of (6). Then we introduce an objective function that takes measurement noise into account and does not cause computational overhead. Finally, we apply this formulation to a voxel-based multiview registration algorithm as an application example.

A. On the Optimal Condition of EVM

EVM actually provides an elegant formulation that can be considered as the basis of multiview registration. Both BALM and EF are derived from (3). Therefore the optimality is self-evident. Here we provide theoretical analysis for an interesting finding that Plane Landmark and EVM can actually be unified under certain conditions. As the $\lambda_3(\boldsymbol{\Sigma})$ represents the "thickness" of the aggregated point clouds, our intuition is that if the local features share the same parameters, the "thickness" should be minimized. Following this intuition, the optimal condition is given as:

$$\begin{cases} \mathbf{R}_k \cdot \left(\mathbf{R}_{\boldsymbol{\Sigma}_k}\mathbf{e}_z\right) \parallel \hat{\mathbf{n}}, & \forall k, \\ \left(\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k'}\right) \perp \hat{\mathbf{n}}, & \forall k, k' \end{cases}, \tag{9}$$

where $\mathbf{R}_{\boldsymbol{\Sigma}_k}$ is from the decomposition result: $\boldsymbol{\Sigma} = \mathbf{R}_{\boldsymbol{\Sigma}_k}\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_k}\mathbf{R}_{\boldsymbol{\Sigma}_k}^T$ and $\mathbf{e}_z = [0, 0, 1]^T$.

Next, we prove that this is actually the optimal condition that minimizes the objective function (3). To this end, our goal is to prove: For the minimization problem in (6), a solution $\{\hat{\mathbf{T}}_1, \ldots, \hat{\mathbf{T}}_{n_k}\}$ is optimal iff it satisfies (9). We begin with the proof of sufficiency.

Lemma. 2 (Weyl's inequality) Given $\mathbf{M} = \mathbf{N} + \mathbf{R}$, where $\mathbf{N}$ and $\mathbf{R}$ are $n \times n$ symmetric matrices, the following inequality

Algorithm 1: Voxel-based Multiview Registration.

Input: $\{\mathcal{P}_k\}$, $\{\bar{\boldsymbol{\xi}}_k\}$, r
Result: $\{\hat{\boldsymbol{\xi}}_k\}$, $\hat{\boldsymbol{\pi}}$
// initialize map with resolution r
$\mathcal{M}$ = initializeVoxelMap(r)
for $\mathcal{P}_k$ *in* $\{\mathcal{P}_k\}$ **do**
  castCloudToMap($\mathcal{P}_k$, $\mathcal{M}$)
end
$\mathcal{V}$ = countActiveVoxels($\mathcal{M}$)
for $\mathcal{V}_k$ *in* $\mathcal{V}$ **do**
  checkInliers($\mathcal{V}_k$)
end
while *not converged* **do**
  $\hat{\boldsymbol{\pi}}$ = updateFeatureParam($\hat{\boldsymbol{\xi}}_k$)
  $\mathbf{J}$, $\delta\mathbf{r}$ = computeJacobianAndResidual($\mathcal{V}$, $\hat{\boldsymbol{\pi}}$)
  $\{\hat{\boldsymbol{\xi}}_k\}$ = updateState($\mathbf{J}$, $\delta\mathbf{r}$)
end

holds for $1 \leqslant i \leqslant n$:

$$\lambda_i(\mathbf{N}) + \lambda_n(\mathbf{R}) \leqslant \lambda_i(\mathbf{M}) \leqslant \lambda_i(\mathbf{N}) + \lambda_1(\mathbf{R}).$$

In Section IV-B2, we show that $\boldsymbol{\Sigma} = \sum_k \frac{n_k}{n}(\boldsymbol{\Sigma}_k + \boldsymbol{\Sigma}_{\boldsymbol{\mu}_k})$, where $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mu}_k}$ are both symmetric semi-positive matrices and the detailed derivation can be found in Section IV-B2. Then by applying Weyl's inequality on $\boldsymbol{\Sigma}$, we have

$$\begin{aligned}\lambda_3(\boldsymbol{\Sigma}) &\geqslant \sum_k \frac{n_k}{n}\left(\lambda_3(\boldsymbol{\Sigma}_k) + \lambda_3\left(\boldsymbol{\Sigma}_{\boldsymbol{\mu}_k}\right)\right) \\ &\geqslant \sum_k \frac{n_k}{n}\lambda_3(\boldsymbol{\Sigma}_k). \end{aligned} \tag{10}$$

This inequality gives a lower bound of $\lambda_3(\boldsymbol{\Sigma})$. Next we show that if (9) is satisfied, the energy function reaches the lower bound.

Theorem: 1 If the optimal conditions (9) are satisfied, the following equality holds:

$$\boldsymbol{\Sigma}\hat{\mathbf{n}} = \left(\sum_k \frac{n_k}{n}\lambda_3(\boldsymbol{\Sigma}_k)\right)\hat{\mathbf{n}}. \tag{11}$$

Proof: By multiplying $\hat{\mathbf{n}}$ to both sides, we have

$$\begin{aligned}\boldsymbol{\Sigma}\hat{\mathbf{n}} &= \sum_k \frac{n_k}{n}\left(\boldsymbol{\Sigma}_k\hat{\mathbf{n}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu}_k}\hat{\mathbf{n}}\right) = \sum_k \frac{n_k}{n}\left(\boldsymbol{\Sigma}_k\hat{\mathbf{n}}\right) \\ &= \sum_k \frac{n_k}{n}\left(\boldsymbol{\Sigma}_k\mathbf{R}_k \cdot (\mathbf{R}_{\boldsymbol{\Sigma}_k}\mathbf{e}_z)\right) \quad \text{(collinerity)} \\ &= \sum_k \frac{n_k}{n}\left(\mathbf{R}_k\mathbf{R}_{\boldsymbol{\Sigma}_k}\boldsymbol{\Lambda}_k \underbrace{\mathbf{R}_{\boldsymbol{\Sigma}_k}^T\mathbf{R}_k^T\mathbf{R}_k\mathbf{R}_{\boldsymbol{\Sigma}_k}}_{\mathbf{I}_{3\times 3}}\mathbf{e}_z\right) \\ &= \sum_k \frac{n_k}{n}\lambda_3(\boldsymbol{\Sigma}_k)\mathbf{R}_k\mathbf{R}_{\boldsymbol{\Sigma}_k}\mathbf{e}_z \\ &= \left(\sum_k \frac{n_k}{n}\lambda_3(\boldsymbol{\Sigma}_k)\right)\hat{\mathbf{n}}. \end{aligned} \tag{12}$$

Based on Theorem 1, $\sum_k \frac{n_k}{n}\lambda_3(\boldsymbol{\Sigma}_k)$ is the eigenvalue of $\boldsymbol{\Sigma}$ and its corresponding eigenvector is $\hat{\mathbf{n}}$. In addition, with the conditions satisfied, the lower bound of $\lambda_3(\boldsymbol{\Sigma})$ is reached. This shows that (9) is sufficient for the optimality. Then, we raise a counter example to prove the necessity.

Definition: 1 (Rayleigh quotient) Given a symmetric matrix $\mathbf{A}$, the Rayleigh quotient is defined by:

$$R(\mathbf{A}, \mathbf{x}) = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}. \tag{13}$$

Lemma: 3 (Bounds of Rayleigh quotient) For a 3×3 symmetric matrix $\mathbf{A}$ and vector $\mathbf{x}$, the bounds of Rayleigh quotient are given by:

$$\lambda_3(\mathbf{A}) \leqslant R(\mathbf{A}, \mathbf{x}) \leqslant \lambda_1(\mathbf{A}) \tag{14}$$

Suppose we have a solution that does not satisfy (9), we show that we can always achieve a lower value of Rayleigh quotient by perturbing the motion parameters. Given an optimal solution $\hat{\mathbf{x}}$, and there exists $\mathbf{R}_k\mathbf{R}_{\Sigma_k}\mathbf{e}_z \not\perp$. We perturb $\mathbf{R}_k$ by $\mathbf{R}'_k = \mathbf{R}_k\delta\mathbf{R}_k$, so that we have $\mathbf{R}'_k\mathbf{R}_{\boldsymbol{\Sigma}_k}\mathbf{e}_z \perp \hat{\mathbf{n}}$. Accordingly, we show that the updated Rayleigh quotient is given as:

$$\hat{\mathbf{n}}^T\boldsymbol{\Sigma}\hat{\mathbf{n}} = \frac{n_p}{n}\hat{\mathbf{n}}^T\boldsymbol{\Sigma}_k\hat{\mathbf{n}} + c > \frac{n_p}{n}\hat{\mathbf{n}}^T\boldsymbol{\Sigma}_p\left(\mathbf{R}'_p\right)\hat{\mathbf{n}} + c,$$

which does not reach the lower bound. Recall *Lemma. 3*, $\hat{\mathbf{n}}$ is not the optimal solution, thus the necessity is proved.

Remark 4: The above proof implicitly unifies the theory of PL-based and EVM-based methods. However, two factors position PL-based methods against other solutions. The first is that the above proof assumes $\mathbf{n}$ is optimal under current estimation. For PL-based methods, as the plane parameter is optimized based on local parameterization, there is no guarantee that it is the optimal solution. The second issue is that certain information loss is encountered, especially with the measurement noise. To handle the measurement noise properly, we re-formulate the equation with a weighting scheme described in the next section.

B. Implementation on Multiview Registration

1) Objective Function: Starting from the above proof, we now provide the proposed objective function based on the idea of both PL-based and EVM-based methods, which is given by:

$$\begin{aligned}\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} &\sum_{k=1}^{N} n_k\lambda_1(\boldsymbol{\Sigma}_k)\|\mathbf{R}_k\mathbf{R}_{\boldsymbol{\Sigma}_k}\mathbf{e}_x \cdot \mathbf{n}\|^2 \\ &+ \sum_{k=1}^{N} n_k\lambda_2(\boldsymbol{\Sigma}_k)\|\mathbf{R}_k\mathbf{R}_{\boldsymbol{\Sigma}_k}\mathbf{e}_y \cdot \mathbf{n}\|^2 \\ &+ \sum_{k=1}^{N} n_k\|\mathbf{n}^T(\mathbf{R}_k\boldsymbol{\mu}_k + \mathbf{t}_k - \boldsymbol{\mu})\|^2, \end{aligned} \tag{15}$$

where $\mathbf{R}_{\boldsymbol{\Sigma}_k}$ is decomposed in advance of the optimization. In the formulation, the first two terms constrain the rotational component of individual poses, and the last term constrains both. This gives a more clear interpretation geometrically. In the implementation, $\mathbf{n}$ is supposed to be optimal under the current estimation, and this is actually not included in the optimization update. After

each step, we re-calculate $\mathbf{n}$ with aggregated covariance $\mathbf{\Sigma}$. This formulation differs from the previous methods that 1) compared to Landmark-based methods, all the measurements are taken into account along with the noise (e.g., λ_i); 2) the residuals are naturally formed in a least-squares fashion, allowing efficient Hessian matrix approximation; 3) without loss of generality compared to the Homogeneous representation; 4) the objective function is only related to frame poses, thus there is no need for point-wise computation or downsampling the input point-cloud.

2) Local Distribution Aggregation: Another major concern of previous work is mentioned in *Remark. 1*, i.e., the computational cost of feature parameter updating. When pose update is performed, the update of sample mean and covariance require significant computation. Here we derive a close-formed solution for the $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. Moreover, the time complexity is then *independent* of the total number of points $\sum_k n_k$ and only dependent on the number of frames N.

We can first estimate $\boldsymbol{\mu}_k^\ell$ and $\mathbf{\Sigma}_k^\ell$ under $\mathcal{F}_k$. Given the corresponding pose of $\mathcal{F}_k$, the sample mean and covariance under $\mathcal{W}$ can be derived by linear transformation:

$$\begin{cases} \boldsymbol{\mu}_k = \mathbf{R}_k \boldsymbol{\mu}_k^\ell + \mathbf{t}_k \\ \mathbf{\Sigma}_k = \mathbf{R}_k \mathbf{\Sigma}_k^\ell \mathbf{R}_k^T \end{cases}, \tag{16}$$

With the estimation results of different subsets, the distribution for aggregated point cloud can be derived in closed form, given by:

$$\begin{cases} \boldsymbol{\mu} = \sum_k \frac{n_k}{n} \boldsymbol{\mu}_k \\ \mathbf{\Sigma} = \sum_k \frac{n_k}{n} \left(\mathbf{\Sigma}_k + \mathbf{\Sigma}_{\boldsymbol{\mu}_k} \right) \end{cases}, \tag{17}$$

where the second term $\mathbf{\Sigma}_{\boldsymbol{\mu}_k}$ is given by:

$$\mathbf{\Sigma}_{\boldsymbol{\mu}_k} = (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T. \tag{18}$$

For the simplicity, the detailed derivation is provided in [26]. With this formulation, the update of feature parameters can be efficient, as the point-wise update is avoided.

3) Voxel-Based Association: To tackle a specific multiview registration problem, we implement a voxel-based multiview registration pipeline, which additionally deals with association and feature selection. The detailed implementation of this system is illustrated in the algorithm. 1. In the beginning, we build up a voxel map where the voxels are stored in a hash table with unique indices. The voxel map is established with a specific resolution r. Given each frame's point-cloud data, we then cast each frame into the voxel map with the initial frame pose. We assume that each voxel corresponds to a global feature, and scanned points inside this voxel of a specific frame are considered the local observation of this feature. The correspondences across different frames are associated according to this voxel representation. Before the optimization, we determine active voxels and inlier local observations by:

1) the number of local measurements is sufficient;
2) the ratio between eigenvalues is appropriate.

We aggregate the local distributions in each voxel with the updated frame poses at each optimization step, and then we estimate the feature parameters with the aggregated covariance $\mathbf{\Sigma}$. After that, we first compute Jacobians and residuals, and apply the LM method to calculate parameter update $\delta\mathbf{x}$. We iterate until the optimization converges, and finally, the optimal poses for different frames are calculated.

V. Experimental Results and Discussions

To evaluate the feasibility and effectiveness of the proposed method, we perform: (1) evaluation on different BA methods on simulation. (2) registration experiment on real-world scenarios.

A. Monte Carlo Simulation

To validate the proposed method against other formulations, we perform extensive Monte Carlo simulations on different methods for multiview registration.

1) Simulation Setup: We follow [12] to design a simulator that randomly generated poses and planar features. For each feature, the local observations in individual frames are also produced by the simulator, as a consequence of which the associations are pre-defined. Besides, to show the advantages and disadvantages of different methods, we validate the performance of different methods with different parameter configurations, including the number of planes (#planes), number of poses (#poses), and noise of the observations (σ); Given a group of observations generated with pre-set parameters,

For the comparison, we use an EVM-based approach EigenFactor (EF), and PL-based method LIPS as baseline methods. Additionally, we re-formulate the constraint proposed by EF into a least-squares manner, which can then be solved using the second-order method (e.g., LM) for better convergence speed. With Cholesky decomposition, we have $\mathbf{S}_i = \mathbf{L}_i \mathbf{L}_i^T$. Then (8) can be rewritten as a least-squares problem, given as:

$$E = \sum_k \boldsymbol{\eta}^T \mathbf{T}_k \mathbf{L}_k \mathbf{L}_k^T \mathbf{T}_k^T \boldsymbol{\eta} = \sum_k \left\| \mathbf{L}_k^T \mathbf{T}_k^T \boldsymbol{\eta} \right\|_2^2.$$

With this least-squares formulation, the Jacobian computation can be much efficient and the second-order methods (e.g., GN or LM) provides better convergence speed. We denote this baseline as **EF(LM)**.

2) Results and Discussions: The evaluation is illustrated in Fig. 1. Generally, we observe that compared with other methods, the proposed formulation achieves the best performance considering both accuracy and speed. For EF, we observe that, as stated in [13], the first-order gradient method used by EigenFactor is inefficient to converge. On the contrary, the second-order method is more efficient to converge.

Also, as LIPS does not consider the feature parameter estimation noises, we observe that the corrupted coefficient leads to inaccurate estimation results. On the contrary, EF, EF(LM) and ours estimate the model coefficients after each iteration, which guarantees the coefficients to be optimal under current pose estimation. In other words, such estimation results minimize the objective function with the current pose estimation results. LIPS, however, performs parameterization on local observations, which is considered to lose certain information in the raw measurements. This is efficient in some test cases. However, as the model coefficients are estimated from local observations, it is not guaranteed to be optimal.

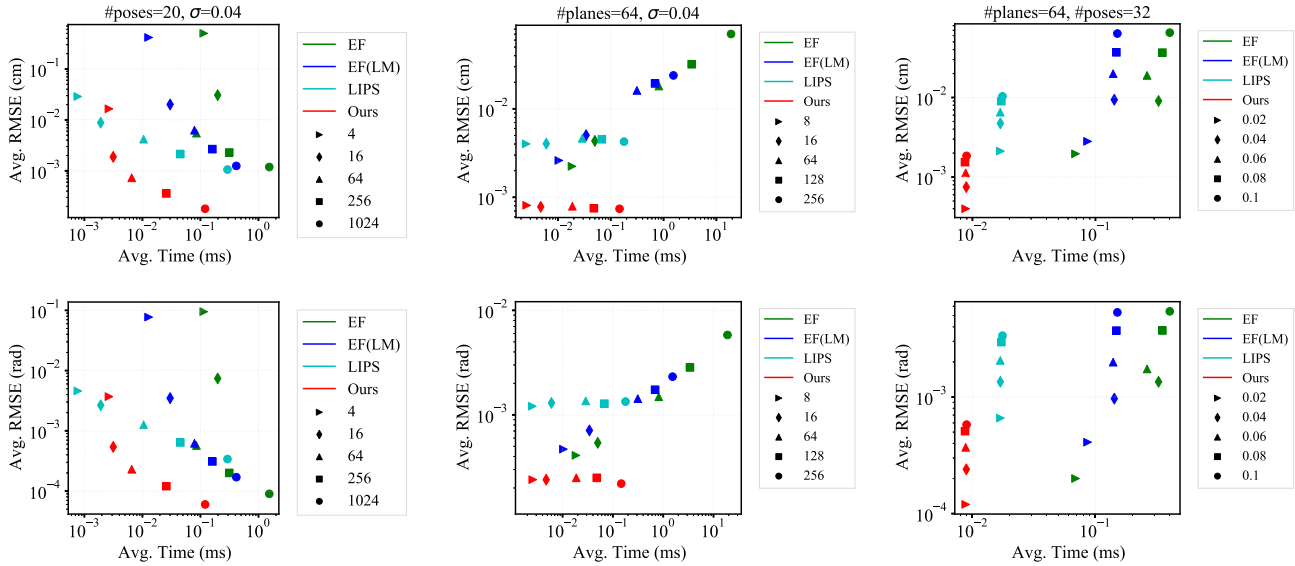

Fig. 1. Speed-accuracy tests in simulation for different BA methods, where the y-axis is the corresponding translational (Top Row) and rotational (Bottom Row) error, with respect to different number of landmarks (Left Column), number of poses (Middle Column), and noise level (Right Column).

TABLE I
EVALUATION OF REGISTRATION PERFORMANCE ON ETHZ REGISTRATION DATASET. WE REPORT AVERAGED RELATIVE POSE ERROR (RPE [CM]) AND ABSOLUTE POSE ERROR (APE [CM]) FOR DIFFERENT METHODS. THE **BEST** AND *SECOND BEST* RESULTS ARE HIGHLIGHTED

Method		Apart.	Haupt.	Stairs	Moun.	Gaze.S.	Gaze.W.	Wood.S	Wood.A	**Average**
Global	TEASER	5.5 / 15.1	2.7 / 10.2	4.3 / 10.0	6.9 / 36.6	3.4 / 8.0	2.8 / 5.5	4.2 / 7.3	3.2 / 8.8	4.1 / 12.7
	FGR	4.9 / 22.0	8.7 / 86.4	4.6 / 6.2	8.6 / 30.3	3.8 / 7.3	4.5 / 12.7	3.6 / 10.8	3.3 / 7.0	5.2 / 22.8
Global+Sync	TEASER	4.1 / 10.7	2.0 / 3.1	3.7 / 5.6	6.8 / 13.7	3.2 / 3.4	2.8 / 2.3	5.1 / 4.6	3.6 / 4.7	3.9 / 6.0
	FGR	5.4 / 17.5	- / -	4.3 / 8.1	10.4 / 28.7	4.1 / 6.2	6.0 / 9.8	5.5 / 19.0	4.3 / 21.4	6.5 / 27.2
F2F	ICP(pt2pt)	2.1 / 9.4	1.5 / 7.0	1.5 / 3.9	3.4 / 26.4	1.4 / 6.6	1.1 / 5.5	2.3 / 10.0	1.9 / 14.9	1.9 / 10.5
	ICP(pt2pl)	1.2 / 3.9	**0.4** / 1.5	*0.8* / 1.8	**2.1** / 14.8	1.1 / 4.4	*0.9* / 5.1	2.5 / 9.3	1.8 / 13.4	*1.3* / 6.8
	GICP	- / -	*0.5* / 1.2	**0.7** / 1.7	4.8 / 45.1	4.8 / 57.2	2.3 / 16.9	2.6 / 9.7	2.0 / 13.0	3.1 / 47.8
	VGICP	**0.6** / **1.1**	**0.4** / **0.6**	0.9 / 1.6	35.5 / 84.3	11.0 / 37.7	- / -	3.0 / 7.7	1.8 / 8.5	10.9 / 49.2
	NDT	3.1 / 11.5	2.5 / 6.2	3.3 / 8.0	4.3 / 21.3	3.2 / 6.3	3.1 / 9.0	3.8 / 13.2	3.4 / 20.1	3.3 / 12.0
F2M	ICP(pt2pt)	1.8 / 4.5	1.1 / 2.0	1.5 / 3.5	3.8 / 23.2	*1.0* / *1.8*	**0.7** / *1.3*	**1.8** / *2.6*	**1.1** / 4.9	1.6 / 5.5
	ICP(pt2pl)	3.1 / 30.2	*0.5* / **0.6**	**2.1** / 4.4	**2.1** / *8.6*	1.7 / 2.2	**0.7** / 1.4	2.4 / 3.2	1.5 / 4.7	1.8 / 6.9
	GICP	3.9 / 35.5	0.6 / **0.6**	*0.8* / **1.3**	4.2 / 21.6	1.1 / 2.7	**0.7** / *1.3*	2.1 / 2.5	1.5 / *4.3*	1.8 / 8.7
	VGICP	- / -	1.2 / 2.4	16.1 / 24.5	- / -	2.1 / 2.0	1.8 / 1.5	11.8 / 11.3	- / -	33.2 / 92.2
	NDT	3.8 / 5.0	1.2 / 2.5	3.6 / 3.9	5.0 / 12.3	2.5 / 3.3	2.1 / 2.3	3.9 / 3.7	3.4 / 4.5	3.2 / 4.7
BA	BALM	2.4 / 3.8	0.6 / *0.7*	1.1 / 1.7	4.0 / 11.1	1.7 / 2.5	1.2 / 2.0	2.8 / 3.2	2.1 / 4.8	2.0 / *3.7*
	Ours	*1.1* / *1.8*	**0.4** / 1.0	0.9 / *1.4*	*2.4* / **5.8**	**0.8** / **1.0**	**0.7** / **0.8**	*2.0* / **1.9**	*1.2* / **1.9**	**1.2** / **2.0**

Besides, while LIPS is efficient when the number of features is small, with the number of features growing, LIPS becomes less efficient. Considering a typical range-based registration problem, there are generally thousands of features, which would introduce a significant number of parameters to optimize.

B. Real World Experiemnts

We perform experiments on the ETHZ Registration Dataset [27], a widely used dataset for registration covering both structured and unstructured scenarios. Each sequence contains 30∼45 frames of point cloud scans. For the details of the dataset, we refer the reader to [27]. To validate our method, we evaluate methods in several categories for comparison, which are described as follows: **Global** registration methods TEASER [28] and FGR [29], which register consequential frames without initial guess; **Global+Sync** methods, which extend global methods to utilize more inter-frame information. In the implementation, pairs of point cloud that share an overlap ratio over 60% are registered by a global method. Then, each method fine-tunes the initial pairwise estimations using standard Pose Graph Optimization (PGO); **Frame-To-Frame** (F2F) methods that contain several ICP variants, which are most widely used for point cloud registration, including point-to-point ICP (ICP(pt2pt)), point-to-plane ICP (ICP(pt2pl)), NDT, GICP and VGICP; **Frame-To-Map** (F2M) methods extending basic registration methods in a frame-to-map fashion, which better utilize inter-frame information. In the implementation, each scan is registered with first the previous scan and then the global map. After registration, the transformed scan is integrated to

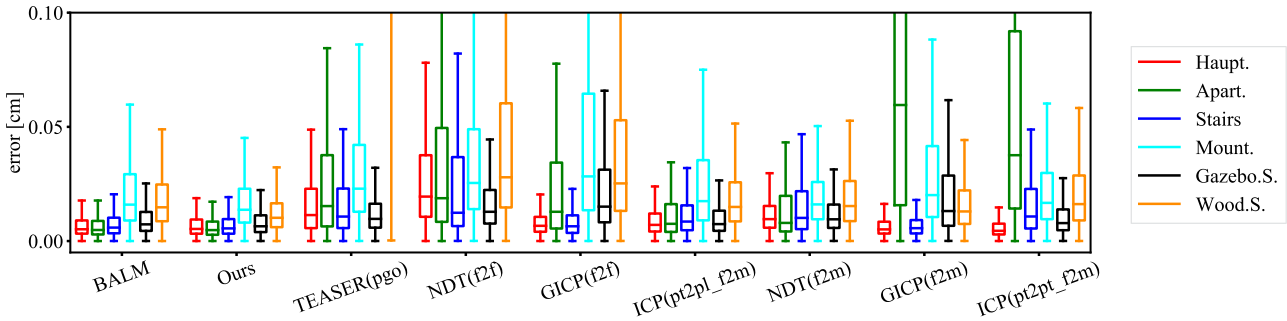

Fig. 2. Evaluation on the structural accuracy of the reconstructed point cloud against the groundtruth.

the global map; **BA** methods including BALM and our method (Ours), which simultaneously optimize poses of several frames.

For F2F, F2M and BA methods, the estimations from TEASER are used as the initial guess. We found that F2F and F2M methods occasionally fail when using raw measurements directly. Therefore, for these methods, the input point cloud is downsampled with a resolution of 0.1 m.

1) Evaluation of Registration Accuracy: We evaluate the accuracy of the registered frame poses. The evaluation metrics are the Relative Pose Error(RPE) and the Absolute Pose Error(APE) [30]. RPEs are computed by the estimated poses of two adjacent frames, while the APEs are calculated between the estimated trajectory and the groundtruth after translational alignment.

Table I illustrates the evaluation of the registration accuracy. Generally, by exploiting inter-frame constraints and jointly optimize the pose parameters, the average registration error of BA methods is on par with or better than other category of methods. Our method achieves the best RPE (1.2 cm) and APE (2.0 cm) while BALM achieves the second best APE (3.7 cm). The RPEs of different methods are actually very close and there is only trivial or no improvement using BA. For example, on sequence Wood.A., the RPE of ICP(pt2pl) and ours are 1.1 cm and 1.2 cm, respectively. However, for the APE which represent the global consistency, our method is generally better than other baselines. NDT(F2M) and VGICP(F2M) is very similar pipeline compared to ours in implementation. From the results, we observe that while NDT(F2M) and VGICP(F2M) have accumulated drifts and inevitable estimation inaccuracy (even fails in some cases), our method well maintains the global consistency in all the sequences. This verifies the benefit of using BA in multiview registration.

2) Evaluation of the Reconstruction Quality: We concatenate the point cloud with the estimated poses for a global map, and then align it with the 3D model from the groundtruth. The structural error is evaluated by calculating the distance between each point and its closest neighbor from the groundtruth geometry. Fig. 2 shows the distribution of structural error on 6 representative sequences. The results show that our reconstruction is on par with or better than the existing method. The structural error is generally caused by the drift in pose estimation, therefore these qualitative results also confirm the quantitative evaluation in Table I. Fig. 3 visualizes the reconstruction quality

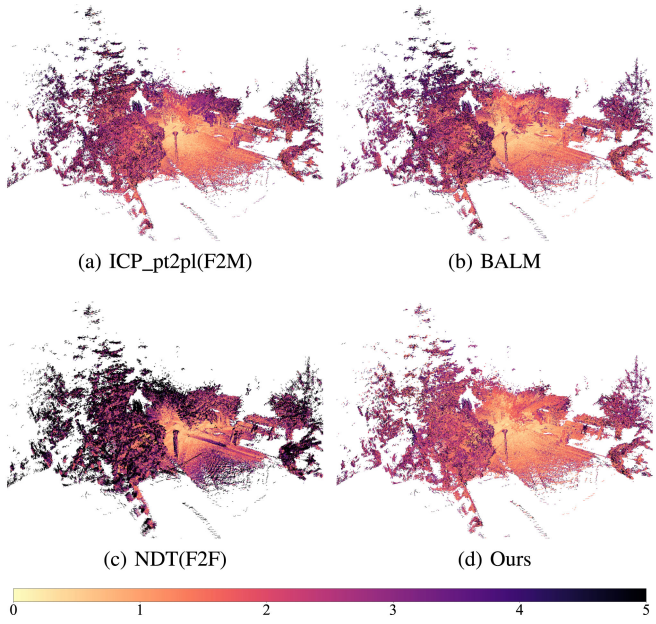

Fig. 3. Heat maps showing the reconstruction error of 4 methods from a fixed view on sequence Gazebo.S. The range of error is set to 0-5 cm as showing in the bottom colorbar.

of our method and 3 other representative methods on Gazebo.S. sequence. Despite the unstructured environment, our method works well and the reconstructed point cloud well aligns with the ground truth. In addition, we observe that with our formulation, the regions with large structural error are generally non-planar cases. This indicates that more generic model would contribute to the registration performance.

3) Comparison With BALM: We further perform detailed comparison on registration accuracy and runtime with BALM, and the results are shown in Fig. 4. We adopt two variants of our method, denoted as Ours(20) and Ours(inf), where the number of optimization iterations are set to 20 and infinite (iterating until convergence), respectively. With the variances of input resolution, the position error of each method is close and the runtime of our methods does not change significantly. In contrast, the runtime of BALM increase a lot when directly using the raw measurements. This is consistent with our observation in Section III that the complexity of BALM is dependent of

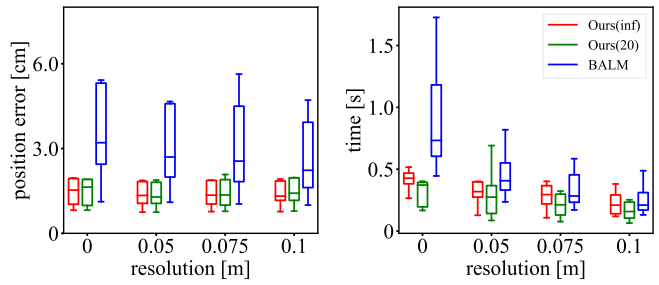

Fig. 4. Evaluation on the translational error and runtime w.r.t. the resolution of the input point cloud.

the number of input point cloud, while that of our method is dependent of the number of features. Originally, we expect the runtime of our method is constant when the resolution changes. However, in the experiment, we found that the changes affect the association to some extend, as a consequence of which the runtime of our method also increases if the resolution of input data is high.

VI. Conclusion

In this work, we have reviewed prior arts on the problem of multiview registration in detail, especially for PL-based and EVM-based methods. To introduce our formulation on this problem, we have first provided a theoretical analysis on the EVM-based formulation's optimal condition, yielding that it can be uniformed with PL-based methods in a noise-less situation. Then, we have introduced a different objective function that weighs rotational, and translational terms by the eigenvalues from decomposition to handle the measurement noise and the computational cost properly. Finally, we have proposed a multiview registration system that utilizes the above formulation, voxel-based data management for feature association and local distribution aggregation for optimal state calculation. Both simulation and the real-world experimental results validate the proposed method.

Our current implementation focuses on multiview point cloud registration, which leaves SLAM problem with sparse LiDAR scans unexplored. We consider that it would be meaningful to apply the proposed method in a range-based SLAM system and perform further analysis. In the future, we would like to investigate more effective approaches in data quantization further. Unlike some KD-Tree-based methods, the association step relies on the initial estimation in the current voxel-based implementation. Although it is highly efficient, it is supposed to be more sensitive to the local minimum. This is significant to the feature association for state estimation.

References

[1] T. Liu *et al.*, "Hercules: An autonomous logistic vehicle for contactless goods transportation during the COVID-19 outbreak," 2020, *arXiv:2004.07480*.

[2] H. Huang, H. Ye, Y. Sun, and M. Liu, "GMMLoc: Structure consistent visual localization with gaussian mixture models," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 5043–5050, Oct. 2020.

[3] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau, "Towards a general multi-view registration technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 5, pp. 540–547, May 1996.

[4] O. D. Faugeras and M. Hebert, "The representation, recognition, and locating of 3-D objects," *Int. J. Robot. Res.*, vol. 5, no. 3, pp. 27–52, 1986.

[5] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Robot.: Sci. Syst.*, vol. 2, no. 9, 2014.

[6] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "SE-Sync: A certifiably correct algorithm for synchronization over the special euclidean group," *Int. J. Robot. Res.*, vol. 38, no. 2-3, pp. 95–125, 2019.

[7] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms Data Structures*, vol. 1611. International Society for Optics and Photonics, 1992, pp. 586–606.

[8] K. Pulli, "Multiview registration for large data sets," in *Proc. 2nd Int. Conf. 3-D Digit. Imag. Model.*, 1999, pp. 160–168.

[9] D. F. Huber and M. Hebert, "Fully automatic registration of multiple 3D data sets," *Image Vis. Comput.*, vol. 21, no. 7, pp. 637–650, 2003.

[10] J. Weingarten and R. Siegwart, "3D SLAM Using Planar Segments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2006, pp. 3062–3067.

[11] P. Geneva, K. Eckenhoff, Y. Yang, and G. Huang, "LIPS: Lidar-inertial 3D plane slam," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 123–130.

[12] G. Ferrer, "Eigen-factors: Plane estimation for multi-frame and time-continuous point cloud alignment," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 1278–1284.

[13] Z. Liu and F. Zhang, " Balm: Bundle adjustment for lidar mapping," 2020, *arXiv:2010.08215*.

[14] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proc. 3rd Int. Conf. 3-D Digit. Imag. Model.*, 2001, pp. 145–152.

[15] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing ICP variants on real-world data sets," *Auton. Robots*, vol. 34, no. 3, pp. 133–148, Feb. 2013.

[16] T. Shan and B. Englot, "LEGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 4758–4765.

[17] H. Ye, Y. Chen, and M. Liu, "Tightly coupled 3D lidar inertial odometry and mapping," in *Proc. Int. Conf. Robot. Automat.*, 2019, pp. 3144–3150.

[18] J. Jiao, H. Ye, Y. Zhu, and M. Liu, "Robust odometry and mapping for multi-LiDAR systems with online extrinsic calibration," *IEEE Trans. Robot.*, 2021.

[19] Y. Zhu, B. Xue, L. Zheng, H. Huang, M. Liu, and R. Fan, "Real-time, environmentally-robust 3D lidar localization," in *Proc. IEEE Int. Conf. Imag. Syst. Techn.*, 2019, pp. 1–6.

[20] D. Rozenberszki and A. L. Majdik, "LOL: Lidar-only odometry and localization in 3D point-cloud maps," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 4379–4385.

[21] C. Qin, H. Ye, C. E. Pranata, J. Han, S. Zhang, and M. Liu, "LINS: A lidar-inertial state estimator for robust and efficient navigation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 8899–8906.

[22] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Auton. robots*, vol. 4, no. 4, pp. 333–349, 1997.

[23] E. Mendes, P. Koch, and S. Lacroix, "ICP-based pose-graph slam," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, 2016, pp. 195–200.

[24] X. Huang, Z. Liang, X. Zhou, Y. Xie, L. J. Guibas, and Q. Huang, "Learning transformation synchronization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8082–8091.

[25] M. Kaess, "Simultaneous localization and mapping with infinite planes," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 4605–4611.

[26] H. Huang and M. Liu, "Supplemental Materials," [Online]. Available: https://hyhuang1995.github.io/bareg/

[27] F. Pomerleau, M. Liu, F. Colas, and R. Siegwart, "Challenging data sets for point-cloud registration algorithms," *Int. J. Robot. Res.*, vol. 31, no. 14, pp. 1705–1711, Dec. 2012.

[28] H. Yang, J. Shi, and L. Carlone, "TEASER: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, Apr. 2020.

[29] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 766–782.

[30] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 573–580.