

On Bundle Adjustment for Multiview Point Cloud Registration — Supplemental Materials —

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

I. DEFINITIONS AND PROOFS

A. Proof of Lemma. 1

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

$$R(\mathbf{A}, \mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}. \quad (1)$$

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:

$$\lambda_3(\mathbf{A}) \leq R(\mathbf{A}, \mathbf{x}) \leq \lambda_1(\mathbf{A}). \quad (2)$$

The equality holds when \mathbf{x} is exactly the corresponding eigenvector.

The objective function for minimization is a Rayleigh quotient about Σ and \mathbf{n} , shown as:

$$\begin{aligned} E &= \frac{1}{\sum_k n_k} \sum_{i=1}^{n_k} \|\mathbf{n}^T (\mathbf{R}_k \mathbf{p}_{ki} + \mathbf{t}_k - \boldsymbol{\mu})\|_2^2 \\ &= \mathbf{n}^T \left(\frac{1}{\sum_k n_k} \sum_{i=1}^{n_k} (\mathbf{p}'_{ki} - \boldsymbol{\mu})(\mathbf{p}'_{ki} - \boldsymbol{\mu})^T \right) \mathbf{n} \quad (3) \\ &= \mathbf{n}^T \Sigma \mathbf{n} = \frac{\mathbf{n}^T \Sigma \mathbf{n}}{\mathbf{n}^T \mathbf{n}} = R(\Sigma, \mathbf{n}). \end{aligned}$$

With *Lemma. 3*, we have $E = \lambda_3(\Sigma)$ when $\mathbf{n} = \mathbf{R}_\Sigma \mathbf{e}_3$. We also refer the readers to [1].

$$\begin{aligned} \Sigma &= \frac{1}{n} \sum_{k=1}^N \sum_{i=1}^{n_k} (\mathbf{p}'_{ki} - \boldsymbol{\mu})(\mathbf{p}'_{ki} - \boldsymbol{\mu})^T \\ &= \frac{1}{n} \sum_{k=1}^N \sum_{i=1}^{n_k} (\mathbf{p}'_{ki} - \boldsymbol{\mu}_k + \boldsymbol{\mu}_k - \boldsymbol{\mu})(\mathbf{p}'_{ki} - \boldsymbol{\mu}_k + \boldsymbol{\mu}_k - \boldsymbol{\mu})^T \\ &= \frac{1}{n} \sum_{k=1}^N \left(\underbrace{\sum_{i=1}^{n_k} (\mathbf{p}'_{ki} - \boldsymbol{\mu}_k)(\mathbf{p}'_{ki} - \boldsymbol{\mu}_k)^T}_{n_k \Sigma_k} + \sum_{i=1}^{n_k} \underbrace{\left((\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T + (\mathbf{p}'_{ki} - \boldsymbol{\mu}_k)(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T + (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\mathbf{p}'_{ki} - \boldsymbol{\mu}_k)^T \right)}_{\Sigma_\mu} \right) \\ &= \sum_{k=1}^N \frac{n_k}{n} (\Sigma_k + \Sigma_{\mu_k}) + \frac{1}{n} \sum_{k=1}^N \underbrace{\left(\sum_{i=1}^{n_k} \mathbf{p}'_{ki} - \sum_{i=1}^{n_k} \boldsymbol{\mu}_k \right)}_0 (\boldsymbol{\mu}_k - \boldsymbol{\mu})^T + \frac{1}{n} \sum_{k=1}^N (\boldsymbol{\mu}_k - \boldsymbol{\mu}) \underbrace{\left(\sum_{i=1}^{n_k} \mathbf{p}'_{ki} - \sum_{i=1}^{n_k} \boldsymbol{\mu}_k \right)^T}_0 \\ &= \sum_{k=1}^N \frac{n_k}{n} (\Sigma_k + \Sigma_{\mu_k}) \end{aligned}$$

B. Derivation of Eq. 17

Suppose we have a set of frames $\{\mathcal{F}_k\}_{k=1 \dots N}$. For the same feature, the local observation under \mathcal{F}_k is represented by a set of points $\mathcal{P}_k = \{\mathbf{p}_{k1}, \dots, \mathbf{p}_{kn_k}\}$. The sample mean and covariance for each point set under \mathcal{F}_k are denoted by $\boldsymbol{\mu}_k^\ell$ and Σ_k^ℓ . Given the corresponding pose of \mathcal{F}_k , the sample mean and covariance under \mathcal{W} can be derived by linear transformation:

$$\begin{cases} \boldsymbol{\mu}_k = \mathbf{R}_k \boldsymbol{\mu}_k^\ell + \mathbf{t}_k \\ \Sigma_k = \mathbf{R}_k \Sigma_k^\ell \mathbf{R}_k^T \end{cases}, \quad (4)$$

We show that the distribution for aggregated point cloud can be derived in closed form, given by:

$$\begin{cases} \boldsymbol{\mu} = \sum_{k=1}^N \frac{n_k}{n} \boldsymbol{\mu}_k \\ \Sigma = \sum_{k=1}^N \frac{n_k}{n} (\Sigma_k + \Sigma_{\mu_k}) \end{cases}. \quad (5)$$

where the second term Σ_{μ_k} is given by:

$$\Sigma_{\mu_k} = (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T. \quad (6)$$

The derivation is shown as follows:

$$\begin{aligned} \boldsymbol{\mu} &= \frac{1}{n} \sum_{k=1}^N \sum_{i=1}^{n_k} \mathbf{p}'_{ki} = \frac{1}{n} \sum_{k=1}^N \left(\mathbf{R}_k \sum_{i=1}^{n_k} \mathbf{p}_{ki} + \mathbf{t}_k \right) \\ &= \frac{1}{n} \sum_{k=1}^N (\mathbf{R}_k \boldsymbol{\mu}_k^\ell + \mathbf{t}_k) = \sum_{k=1}^N \frac{n_k}{n} \boldsymbol{\mu}_k, \end{aligned} \quad (7)$$

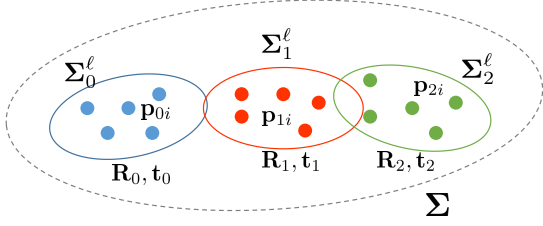


Fig. 1: An illustration of states in the manuscript.

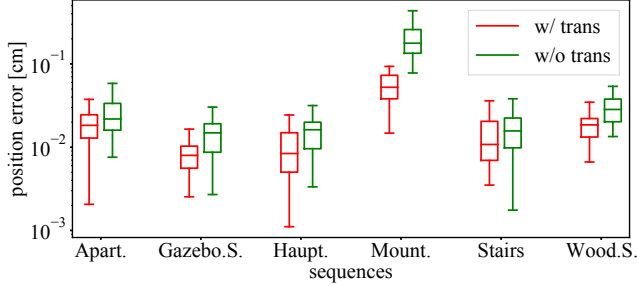


Fig. 2: Comparison of the performance with different error-terms taking into account.

C. State Definitions

An example of state definition is shown in Fig. 1. In the manuscript, we denote the point set under \mathcal{F}_k as \mathcal{P}_k . We assume that points in \mathcal{P}_k belong to an identical feature and shares the same parameters. The sample mean and covariance of \mathcal{P}_k are denoted by μ_k^ℓ and Σ_k^ℓ , respectively. \mathcal{P}_k is subject to a RBT $\mathcal{R}_k, \mathbf{t}_k$, which transform the points from local frame \mathcal{F}_k to the global frame \mathcal{W} . The sample mean and covariance of $\bigcup_k \mathcal{P}_k$ under \mathcal{W} are denoted by μ and Σ .

II. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Parameter Study

In this section, we perform further studies on the impact of voxel size of our implementation. With changing the voxel size, we evaluate the translational and rotational error of our method along with variants of NDT [2] and VGICP [3] in Fig. 3. For different methods, the parameters in the experiments are selected according to this result.

Based on the evaluation, generally, all the methods achieve relatively accurate estimation with voxel size around 0.3–0.4 m. Moreover, we observe that different from the other methods, the error of our methods drops quickly with the increase of voxel size. This could be counter-intuitive because larger voxel size produces sparser data, and sparser data could degrade the registration performance. We conjecture the reason could be that too small voxel size could not guarantee sufficient correspondences to maintain the global consistency. This could be different from the frame-to-frame methods, such as VGICP(nr1). We think it is reasonable that the performance degrades when the voxel size is smaller than 0.2 m. This is because smaller voxel size could not increase the density since the point cloud is already very sparse.

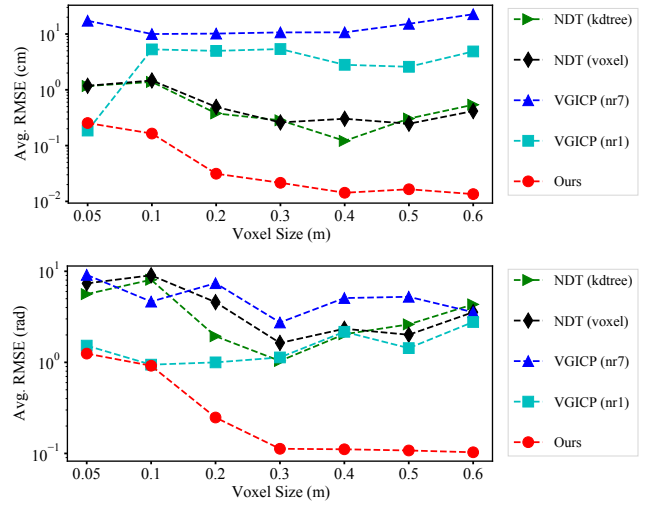


Fig. 3: Evaluation on the accuracy of aligned poses for different methods, when the size of voxel varies from 0.05 to 0.6 m.

B. Ablation Study on the Objective Function

We validate the contributions of different terms in the objective function. We evaluate the registration performance of the original objective function and objective function without taking the translational constraints. The two baselines are denoted as **w/ trans** and **w/o trans**, respectively. As shown in Fig. 2, the results without translational constraints are consistently worse than those with all the terms used in the optimization. This indicates the terms decoupled from the original objective functions do contribute to the registration accuracy.

C. Qualitative Results

To better visualize the reconstructed point cloud from different methods, we provide more qualitative results of more baselines in Fig. 4 and Fig. 5. The results include on sequence Haupt. and Gazebo.S. from ETHZ Registration Dataset [4]. While some traditional methods have certain mis-alignment, methods based on BA generally provide a high-quality reconstruction result.

REFERENCES

- [1] Z. Liu and F. Zhang, “Balm: Bundle adjustment for lidar mapping,” *arXiv preprint arXiv:2010.08215*, 2020.
- [2] M. Magnusson, A. Lilienthal, and T. Duckett, “Scan registration for autonomous mining vehicles using 3d-ndt,” *Journal of Field Robotics*, vol. 24, no. 10, pp. 803–827, 2007.
- [3] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, “Voxelized gicp for fast and accurate 3d point cloud registration,” in *IEEE International Conference on Robotics and Automation (ICRA2021)*, May 2021.
- [4] F. Pomerleau, M. Liu, F. Colas, and R. Siegwart, “Challenging data sets for point-cloud registration algorithms,” *The International Journal of Robotics Research*, vol. 31, no. 14, pp. 1705–1711, Dec. 2012.

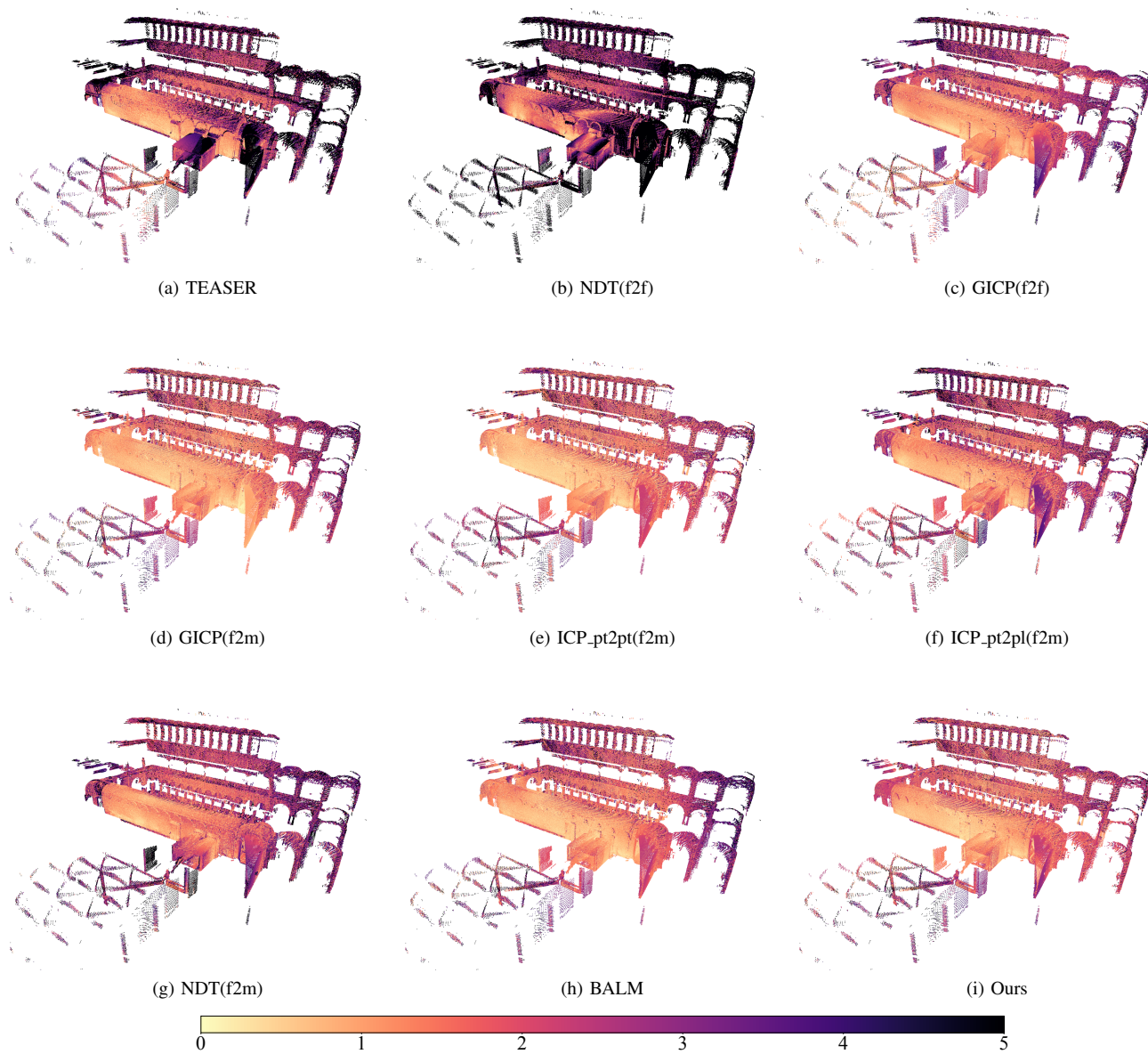


Fig. 4: Heat maps showing the reconstruction error of 4 methods from a fixed view on sequence Haupt. The range of error is set to 0-5 cm as showing in the bottom colorbar.

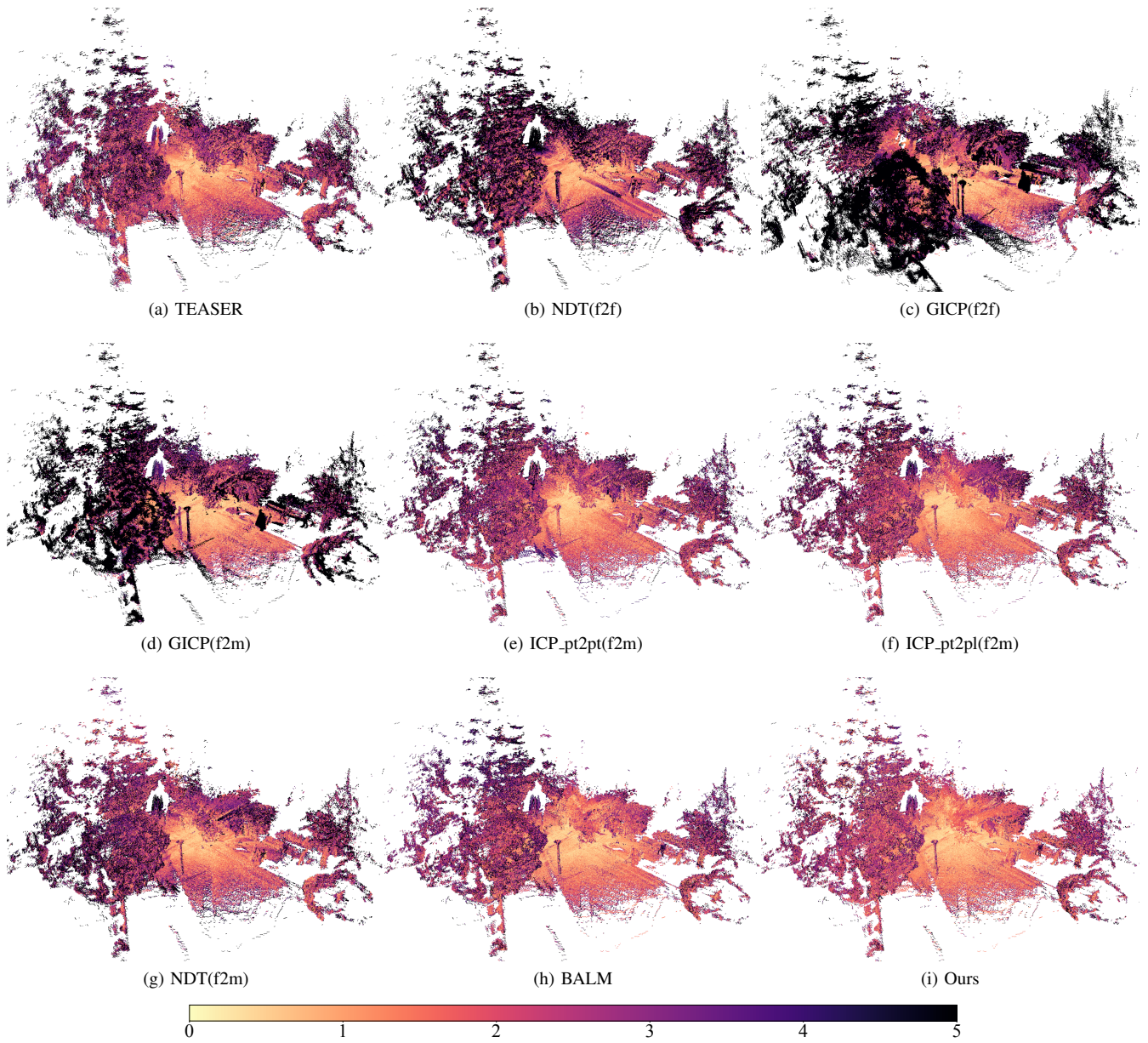


Fig. 5: Heat maps showing the reconstruction error of 4 methods from a fixed view on sequence Gazebo(Summer). The range of error is set to 0-5 cm as showing in the bottom colorbar.