PCR-CG: Point Cloud Registration via Color and Geometry

Yu Zhang[1] Junle Yu[2] Xiaolin Huang[1] Wenhui Zhou[2] Ji Hou[3]

[1]Shanghai Jiaotong University [2]Hangzhou Dianzi University [3]TUM

In Section 1, we introduce the details of the metrics used in the main paper. We additionally show more visual results in Section 2. In Section 3, we analyze the running time of plugging in our proposed module. In Section 4, a robustness analysis is demonstrated to show our proposed module helps stabilize the registration on noisy and sparse inputs.

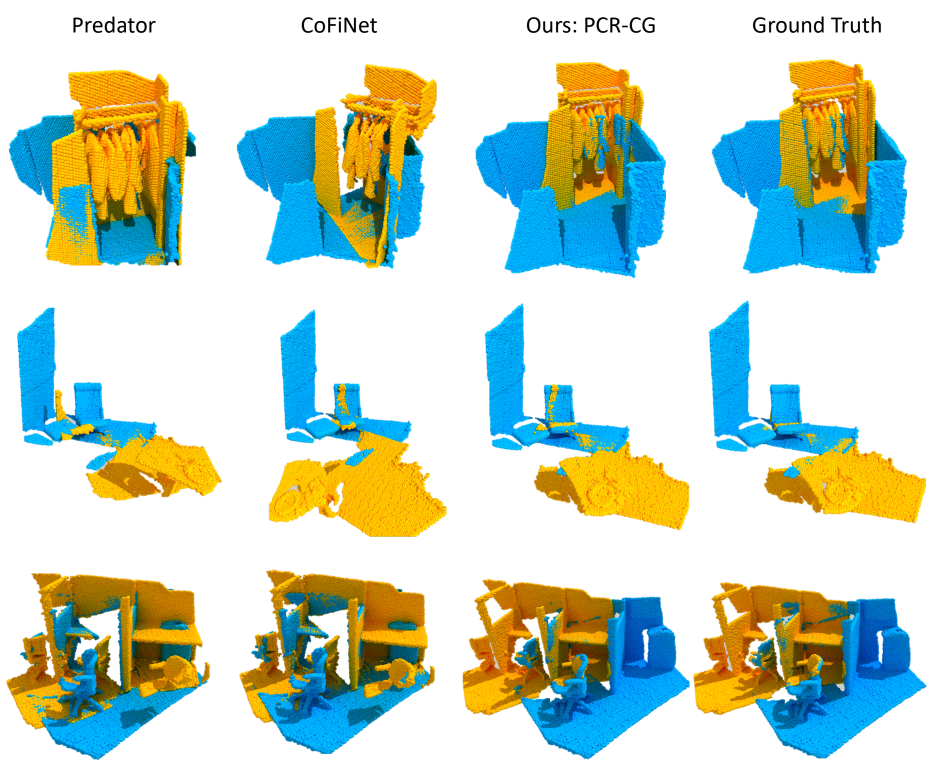

Fig. 1. Our approach outperforms SOTA methods including Predator [4] in those difficult scenarios in 3DLoMatch benchmark.

1 Evaluation Metrics

In this work, we follow the same definitions of metrics as defined in CoFiNet [5] and Predator [4].

Inlier Ratio (IR) measures the fraction of point correspondences $(x_i, y_j) \in \widetilde{\mathbf{C}}$ subject to the Euclidean Norm of residual $\|\overline{\mathbf{T}}_{\mathbf{Y}}^{\mathbf{X}}(x_i) - y_j\|$. $\widetilde{\mathbf{C}}$ denotes the estimated correspondence set. $\overline{\mathbf{T}}_{\mathbf{Y}}^{\mathbf{X}}$ indicates the ground truth transformation between $\mathbf{X}$ and $\mathbf{Y}$. In the metric, we select threshold τ_1=10cm. To this regard, a pair of correspondences count as matched when their Euclidean Norm of residual is smaller than 10cm. Given the estimated correspondence set $\widetilde{\mathbf{C}}$, *Inlier Ratio* of a pair of point clouds $(\mathbf{X}, \mathbf{Y})$ can be calculated by:

$$\mathrm{IR}(\mathbf{X}, \mathbf{Y}) = \frac{1}{|\widetilde{\mathbf{C}}|} \sum_{(x_i, y_j) \in \widetilde{\mathbf{C}}} 1(\|\overline{\mathbf{T}}_{\mathbf{Y}}^{\mathbf{X}}(x_i) - y_j\| < \tau_1), \tag{1}$$

where $1(\cdot)$ represents the indicator function counting the number of correspondences within the threshold τ_1; and $\|\cdot\| = \|\cdot\|_2$ denotes the Euclidean Norm (L_2 distance).

Feature Matching Recall (FMR) measures the fraction of point cloud pairs whose *Inlier Ratio* is larger than a certain threshold $\tau_2 = 5\%$. It is firstly used in [1] and indicates the likelihood of recovering an optimal transformation between two point clouds by a robust pose estimator, e.g., RANSAC [2], based on the predicted correspondence set $\widetilde{\mathbf{C}}$. Given a dataset $\mathcal{D}$ with $|\mathcal{D}|$ point cloud pairs, *Feature Matching Recall* can be computed as:

$$\mathrm{FMR}(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(\mathbf{X}, \mathbf{Y}) \in \mathcal{D}} 1(\mathrm{IR}(\mathbf{X}, \mathbf{Y}) > \tau_2). \tag{2}$$

Registration Recall. Different from aforementioned metrics that measure the quality of extracted correspondences, *Registration Recall* (RR) on the other hand directly measures the performance on the task of point cloud registration. It measures the fraction of point cloud pairs whose Root Mean Square Error (RMSE) is within a certain threshold $\tau_3 = 0.2$m. Given a dataset $\mathcal{D}$ with $|\mathcal{D}|$ point cloud pairs, *Registration Recall* is defined as:

$$\mathrm{RR}(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(\mathbf{X}, \mathbf{Y}) \in \mathcal{D}} 1(\mathrm{RMSE}(\mathbf{X}, \mathbf{Y}) < \tau_3), \tag{3}$$

where for each $(\mathbf{X}, \mathbf{Y}) \in \mathcal{D}$; RMSE of the ground truth correspondence set $\overline{\mathbf{C}}$ is computed as following:

$$\mathrm{RMSE}(\mathbf{X}, \mathbf{Y}) = \sqrt{\frac{1}{|\overline{\mathbf{C}}|} \sum_{(x_i, y_j) \in \overline{\mathbf{C}}} \|\mathbf{T}_{\mathbf{Y}}^{\mathbf{X}}(x_i) - y_j\|^2}. \tag{4}$$

where the estimated transformation $\mathbf{T}_{\mathbf{Y}}^{\mathbf{X}}$ is applied.

Additionally, we follow the original evaluation protocol in 3DMatch [6], which excludes immediately adjacent point clouds with very high overlap ratios.

Relative Translation and Rotation Errors. Given the estimated transformation $\mathbf{T}_{\mathbf{Y}}^{\mathbf{X}} \in SE(3)$ composed of a translation vector $\mathbf{t} \in \mathbb{R}^3$ and a rotation matrix $\mathbf{R} \in SO(3)$. Its Relative Translation Error (RTE) and Relative Rotation Error (RRE) from the ground truth pose $\overline{\mathbf{T}}_{\mathbf{Y}}^{\mathbf{X}}$ are computed as:

$$\text{RTE} = \|\mathbf{t} - \bar{\mathbf{t}}\| \qquad \text{and} \qquad \text{RRE} = \arccos\left(\frac{\text{trace}(\mathbf{R}^\top \overline{\mathbf{R}}) - 1}{2}\right), \tag{5}$$

where $\bar{\mathbf{t}}$ and $\overline{\mathbf{R}}$ are the the ground truth translation and rotation in $\overline{\mathbf{T}}_{\mathbf{Y}}^{\mathbf{X}}$, respectively.

2 More Visualizations

We show more visualizations of our method compared to SOTA methods including Predator [4] in Figure 1. Our approach can significantly improve registration accuracy in most difficult scenarios.

3 Inference Time

Our approach can be easily plugged onto various approaches with a low cost of running time (see Table 1).

	Predator	PCR-CG
Inference time (s)	0.48	0.57

Table 1. Inference time on 5,000 points on 3DLoMatch benchmark. PCR-CG uses two views, Pri3D [3] pre-trained model and the explicit projection.

	Predator [4]	CoFiNet [5]	PCR-CG
Noisy and sparse inputs	56.3	57.8	62.4

Table 2. Registration recall on 3DLoMatch with noisy and sparse inputs. PCR-CG uses two views, Pri3D [3] pre-trained model and the explicit projection.

4 Robustness Analysis

We make input points sparse and noisy by sub-sampling 40% points and randomly jittering 10% points. In Table 2, we empirically demonstrate our module, which embeds deep color features, helps stabilize the registration, which is more robust to sparse and noisy inputs.

References

1. Deng, H., Birdal, T., Ilic, S.: Ppf-foldnet: Unsupervised learning of rotation invariant 3D local descriptors. In: ECCV (2018) 2
2. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981) 2
3. Hou, J., Xie, S., Graham, B., Dai, A., Nießner, M.: Pri3d: Can 3d priors help 2d representation learning? In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5693–5702 (2021) 3
4. Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., Schindler, K.: Predator: Registration of 3d point clouds with low overlap. In: CVPR. pp. 4267–4276 (June 2021) 1, 2, 3
5. Yu, H., Li, F., Saleh, M., Busam, B., Ilic, S.: Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. NeurIPS **34** (2021) 2, 3
6. Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T.: 3DMatch: Learning local geometric descriptors from RGB-D reconstructions. In: CVPR (2017) 3