

Ex-Sim(3)-Reg: 2D-3D Correspondence Pruning via Extended Sim(3) Registration

Pei An^{1*}, Muyao Peng^{1*}, Junfeng Ding¹, Jiaqi Yang^{2†}, and
Liangliang Nan^{3†}

¹ Huazhong University of Science and Technology, China
{anpei96,muyao99,djfenghust}@hust.edu.cn

² Northwestern Polytechnical University, China jqyang@nwpu.edu.cn

³ Delft University of Technology, Netherlands liangliang.nan@tudelft.nl

Abstract. Learning-based image-to-point-cloud (I2P) registration has garnered increasing attention in recent years. Nevertheless, existing methods still struggle with severe outliers under challenging scenarios with unseen, low-inlier, or distorted cases. A fast and robust 2D-3D correspondence pruning method is therefore highly desirable. Recently, a promising scheme lifts 2D-3D correspondences to 3D-3D correspondences using depth priors, casting correspondence pruning as a Sim(3) registration problem. However, depth priors estimated from monocular images are inherently noisy, which undermines the reliability of this scheme. In this paper, to explicitly model non-negligible depth noise, we reformulate correspondence pruning as an extended Sim(3) registration problem and propose a simple yet effective pruning algorithm termed Ex-Sim(3)-Reg. We further provide a theoretical analysis to justify the effectiveness of our method. Extensive experiments on the 7-Scenes, RGBD-V2, ScanNet, and TUM datasets demonstrate that Ex-Sim(3)-Reg achieves up to **24.7% improvement** in registration recall over state-of-the-art baseline methods. Code is released at github.com/anpei96/ex-sim3-demo.

Keywords: Image-to-point-cloud registration · Sim(3) registration · depth prior · 2D-3D correspondence pruning · depth noise

1 Introduction

Image-to-point-cloud (I2P) registration [20] enables a wide range of applications in 3D computer vision, such as visual localization, 3D reconstruction, and multi-sensor calibration [8, 15, 17, 23]. Despite rapid advances in deep learning-based I2P registration [12, 20, 21], current methods still struggle to generalize to unseen, diverse, and complex real-world scenarios [1, 34]. Specifically, they often predict a large number of incorrect 2D-3D correspondences (i.e., outliers), resulting in instability of downstream tasks [4, 14].

* Equal contribution.

† Corresponding author(s).

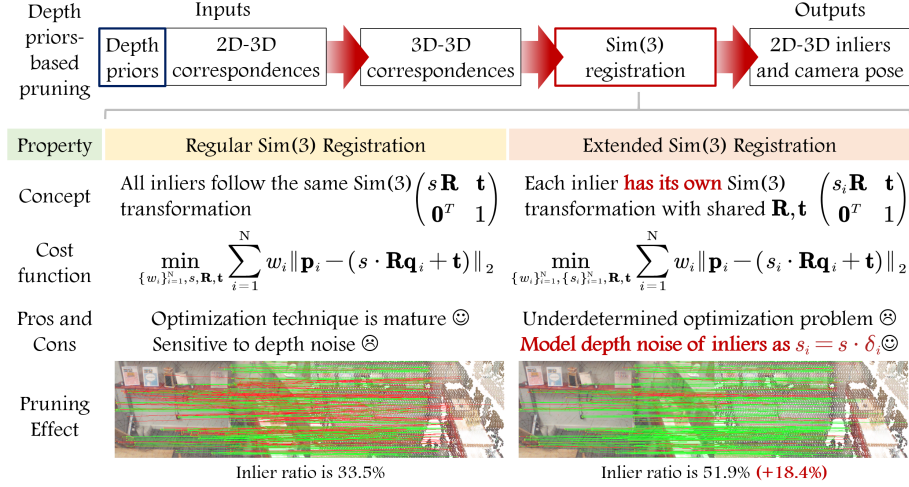


Fig. 1: Motivation. We extend the Sim(3) registration to model the noise δ_i in depth priors as correspondence-specific scale perturbations $s_i = s \cdot \delta_i$. This simple reformulation significantly improves pruning robustness against inaccurate depth priors.

To mitigate outliers, it is crucial to study the 2D-3D correspondence pruning problem. RANSAC-based P3P [24] serves as a default pruning algorithm in most learning-based I2P pipelines [12, 20]. However, when predicted correspondences are heavily contaminated by outliers, RANSAC-based P3P frequently fails to compute an accurate camera pose. Some works formulate 2D-3D correspondence pruning as a non-convex optimization problem [5, 6, 9, 14]. Yet these methods are either computationally expensive [6, 14] or require additional model learning [5, 9], thereby limiting their practicality. Recently, a more practical strategy has emerged that lifts 2D-3D correspondences to 3D-3D correspondences with depth priors [4, 23, 28, 36]. This approach offers two advantages: (i) monocular depth priors can be readily predicted with existing visual foundation models [19, 32]; (ii) metric depth formulates correspondence pruning as a Sim(3) registration problem, which can be effectively solved using classical geometric methods without deep learning [36].

Although depth prior-based pruning is attractive, existing methods rely heavily on highly accurate depth priors [4, 23, 28, 36]. Due to the inherent scale ambiguity of monocular images, depth priors inevitably introduce **non-negligible** errors [22], leading to degraded I2P registration results.

To address this issue, by modeling the depth noise as correspondence-specific scale perturbations (Fig. 1), we reformulate pruning from a regular Sim(3) to an **extended Sim(3)** registration problem. It compensates for depth noise during inlier optimization, which theoretically improves the pruning robustness to the inaccurate depth priors. To solve this underdetermined problem, we develop a heuristic yet effective pruning algorithm named Ex-Sim(3)-Reg. Inliers are identified by searching for maximum cliques in a family of extended compatibility

graphs parameterized by scale hypotheses. Moreover, we theoretically explain the effectiveness of Ex-Sim(3)-Reg by analyzing the structure of extended compatibility graphs. In practice, Ex-Sim(3)-Reg can combine with RANSAC-based P3P [24] for powerful 2D-3D pruning. Results on the 7-Scenes [16], RGBD-V2 [18], ScanNet [13], and TUM [25] datasets demonstrate up to 24.7% gain in registration recall over baseline pruning methods.

Our main contributions are summarized as follows:

- We are **the first** to reformulate depth prior-based 2D-3D correspondence pruning as an extended Sim(3) registration problem that models depth noise as correspondence-specific scale perturbations.
- We propose Ex-Sim(3)-Reg, a fast, non-learning pruning algorithm that approximates the extended Sim(3) registration problem via scale-parameterized compatibility graphs, together with a theoretical analysis of its effectiveness.
- In practice, Ex-Sim(3)-Reg can work with RANSAC-based P3P to achieve better performance than state-of-the-art methods on multiple datasets.

2 Related Works

Up to now, I2P registration typically comprises two procedures: (i) I2P feature matching to generate candidate 2D-3D correspondences, and (ii) 2D-3D correspondence pruning to remove outliers during pose estimation.

I2P feature matching. Current methods for generating I2P correspondences fall into two categories: descriptor-based and flow-based schemes [8]. Descriptor-based methods learn cross-modal features for 2D pixels and 3D points and match them via nearest neighbor search in the feature space [1, 15, 20, 27]. For example, P2-Net [27] uses separate backbones to extract image and point cloud features supervised by a circle loss [26] to improve the correspondence accuracy. Follow-up works improve stability by leveraging patch-level correspondences [20] or by incorporating topological cues [1]. Bie et al. [4] reduces the cross-modal gap by incorporating metric depth predicted from monocular images. Flow-based methods predict dense pixel-wise displacements between an image and a projected point cloud map and are particularly prevalent for outdoor or LiDAR-camera setups [7, 8, 33, 35]. Recent work (i.e., CMRNext [8], 2025) unifies descriptor and flow ideas to boost localization in prior LiDAR maps. In summary, the framework of I2P feature matching has been fully developed in recent years.

2D-3D correspondence pruning. Pruning removes mismatches produced by the matching stage and is critical for robust I2P registration. Early approaches employ global optimization (e.g., branch-and-bound [6]), but these methods are computationally prohibitive for many applications [5]. After 2020, learning-based approaches have been proposed, but most require high initial inlier ratios or additional viewpoint priors [5, 40]. Recently, depth prior-based pruning gained increasing attention, because it allows formulating the original problem as a Sim(3) registration. Bie et al. [4] developed a graph neural network-based method to identify inliers from the Sim(3) registration problem. Murai et al. [23] solved the

Table 1: Comparison of SE(3), Sim(3), and extended Sim(3) registration formulations. In extended Sim(3), each correspondence is associated with an individual scale factor s_i , increasing the number of unknowns from $N+7$ to $2N+6$, which renders the problem underdetermined without additional structural constraints.

Registration problem	Inlier constraint	Objectives	Num. of variables
SE(3)	$\mathbf{p}_i = \mathbf{R}\mathbf{q}_i + \mathbf{t}$	Estimate inliers $\{w_i \in \{0, 1\}\}_{i=1}^N, \mathbf{R}, \mathbf{t}$	$N + 6$
Regular Sim(3)	$\mathbf{p}_i = s\mathbf{R}\mathbf{q}_i + \mathbf{t}$	Estimate inliers $\{w_i \in \{0, 1\}\}_{i=1}^N, \mathbf{R}, \mathbf{t}, s$	$N + 7$
Extended Sim(3)	$\mathbf{p}_i = s_i\mathbf{R}\mathbf{q}_i + \mathbf{t}$	Estimate inliers $\{w_i \in \{0, 1\}\}_{i=1}^N, \mathbf{R}, \mathbf{t}, \{s_i\}_{i=1}^N$	$2N + 6$

Sim(3) registration problem by the iterative closest point (ICP) based optimization, yet they rely on an accurate pose prior to avoid local minima. Yu et al. [36] studied the Sim(3) registration within the depth priors-based two-view geometry problem, which can be extended to correspondence pruning. In summary, despite the progress achieved by existing depth prior-based methods [4, 23, 36], they fail to adequately handle the depth noise during Sim(3) registration. Therefore, it is crucial to develop the novel pruning mechanism that are robust against noisy depth priors.

3 Pruning as an Extended Sim(3) Registration Problem

We motivate the necessity of framing pruning as an extended Sim(3) registration.

3.1 Problem Formulation

Given 2D-3D correspondences $\{\mathbf{c}_i\}_{i=1}^N$ predicted from a learning-based I2P registration model [20], correspondence pruning seeks to filter out outlier matches from $\{\mathbf{c}_i\}_{i=1}^N$. Each correspondence $\mathbf{c}_i \triangleq (\mathbf{l}_i, \mathbf{p}_i)$ consists of a 2D pixel coordinate $\mathbf{l}_i = (u_i, v_i, 1)^T \in \mathbb{R}^3$ and a 3D point coordinate $\mathbf{p}_i = (x_i, y_i, z_i)^T \in \mathbb{R}^3$. A correspondence $\mathbf{c}_i$ is an outlier if it does not satisfy the following constraint:

$$\|\mathbf{l}_i - \pi(\mathbf{p}_i | \mathbf{K}, \mathbf{R}, \mathbf{t})\|_2 \leq \delta_{2d} \quad (1)$$

where $\pi(\cdot)$ is the standard camera projection function [39]. $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ denotes the camera intrinsic matrix. $\mathbf{R} \in \text{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^{3 \times 1}$ represent the camera pose. δ_{2d} is a reprojection error threshold. In essence, 2D-3D correspondence pruning outputs inliers and camera pose.

3.2 Pruning as a Regular Sim(3) Registration

Current methods formulate the pruning as a regular Sim(3) registration problem [4, 23, 36]. Using the pre-trained visual foundation models [31, 32], we estimate a metric depth from a monocular RGB image. 3D-3D correspondence $\mathbf{d}_i \triangleq (\mathbf{q}_i, \mathbf{p}_i)$ is then obtained by lifting $\mathbf{c}_i$ with depth priors. $\mathbf{q}_i = d_i \cdot \pi^{-1}(\mathbf{l}_i | \mathbf{K})$ where $\mathbf{l}_i$ and d_i denote the pixel coordinate and predicted depth of the i -th pixel. $\pi^{-1}(\cdot)$ is the

camera back-projection function [39]. Then, pruning is formulated as a regular Sim(3) registration problem (summarized in Table 1):

$$\min_{\{w_i\}_{i=1}^N, s, \mathbf{R}, \mathbf{t}} \sum_{i=1}^N w_i \|\mathbf{p}_i - (s \cdot \mathbf{R}\mathbf{q}_i + \mathbf{t})\|_2^2 \quad (2)$$

where $s \in \mathbb{R}^+$ is the scale of depth priors. $w_i \in \{0, 1\}$ denotes whether the i -th correspondence is inlier or outlier. In general, Sim(3) registration can be solved within two steps: (i) scale estimation and (ii) pose estimation [30]. The first step targets estimating the scale factor s . By employing the constraint in Eq. 3, the scale s is estimated via the adaptive voting algorithm [30].

$$s = \|\mathbf{p}_i - \mathbf{p}_j\|_2 / \|\mathbf{q}_i - \mathbf{q}_j\|_2, \text{ if } \mathbf{c}_i, \mathbf{c}_j \text{ are inliers.} \quad (3)$$

After scale estimation, Sim(3) registration is reduced to an SE(3) registration. The second step involves estimating $\{w_i\}_{i=1}^N$, $\mathbf{R}$ and $\mathbf{t}$, which can be effectively solved via compatibility graph based SE(3) registration methods [11, 29, 37, 38]. Overall, the regular Sim(3) registration formulated in Eq. 2 can be fully addressed using existing techniques.

3.3 From Regular Sim(3) to Extended Sim(3)

Actually, Eq. 2 exhibits a **critical limitation** in that it does not fully account for the influence of depth noise. Due to scale ambiguity in monocular images, depth priors obtained even from well-trained visual foundation models still suffer from significant errors [22, 36]. In this paper, we model the depth noise by extending the original Sim(3) registration problem. We consider the multiplicative depth noise term δ_i on d_i^3 and derive $\delta_i \mathbf{q}_i = \delta_i d_i \cdot \pi^{-1}(\mathbf{l}_i | \mathbf{K})$. Then, Eq. 2 is extended as a more general registration problem:

$$\min_{\{w_i\}_{i=1}^N, \{s_i\}_{i=1}^N, \mathbf{R}, \mathbf{t}} \sum_{i=1}^N w_i \|\mathbf{p}_i - (s_i \cdot \mathbf{R}\mathbf{q}_i + \mathbf{t})\|_2^2 \quad (4)$$

where $s_i = s \cdot \delta_i$. We call Eq. 4 **extended Sim(3) registration** because every inlier follows a slightly different similarity transformation parametrized by s_i , $\mathbf{R}$, and $\mathbf{t}$. Compared to regular Sim(3) registration, extended Sim(3) registration explicitly optimizes for depth noise, leading to more stable I2P registration.

4 Pruning with Extended Sim(3) Registration

To solve the extended Sim(3) registration, we design a heuristic yet efficient pruning algorithm called **Ex-Sim(3)-Reg** with an in-depth theoretical analysis.

³ multiplicative noise can be equivalently represented as additive noise.

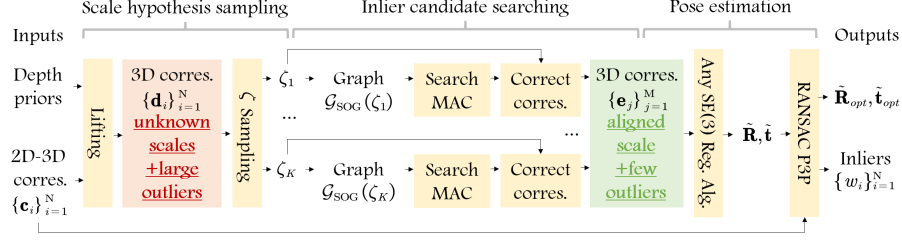


Fig. 2: Pipeline of Ex-Sim(3)-Reg. Through scale hypothesis sampling and inlier candidate search, 3D correspondences $\{\mathbf{d}_i\}_{i=1}^N$ are transformed to $\{\mathbf{e}_j\}_{j=1}^M$. This reduces the extended Sim(3) registration into an SE(3) registration problem, enabling robust inlier estimation using off-the-shelf SE(3) solvers.

4.1 Extended Compatibility Graph

Compatibility graph is a fundamental structure in SE(3) registration [11, 37, 38]. Inliers can be identified by searching the maximum clique in this graph [37]. We extend this concept to solve Eq. 4. Given the lifted 3D-3D correspondences $\{\mathbf{d}_i\}_{i=1}^N$, we define the first-order extended compatibility graph as $\mathcal{G}_{\text{FOG}}(\zeta)$. It consists of N vertices and each vertex denotes a correspondence $\mathbf{d}_i$. ζ is a scale parameter. The adjacency matrix $\mathbf{W}_{\text{FOG}}(\zeta) \in \mathbb{R}^{N \times N}$ describes the edge connections of $\mathcal{G}_{\text{FOG}}(\zeta)$ [37]:

$$\mathbf{W}_{\text{FOG}}(\zeta)_{ij} = \begin{cases} 1, & d_{ij}(\zeta) \leq d_{thr} \\ 0, & d_{ij}(\zeta) > d_{thr} \end{cases}, d_{ij}(\zeta) = \|\mathbf{p}_i - \mathbf{p}_j\|_2 - \|\zeta \cdot (\mathbf{q}_i - \mathbf{q}_j)\|_2 \quad (5)$$

where $d_{thr} \geq 0$ is a distance threshold. For a matrix $\mathbf{W}$, $\mathbf{W}_{ij}$ is the element in the i -th row and j -th column. We further construct the second-ordered extended compatibility graph as $\mathcal{G}_{\text{SOG}}(\zeta)$, which shares the same vertex set as $\mathcal{G}_{\text{FOG}}(\zeta)$ but with a different adjacency matrix $\mathbf{W}_{\text{SOG}}(\zeta)$ [11]:

$$\mathbf{W}_{\text{SOG}}(\zeta) = \mathbf{W}_{\text{FOG}}(\zeta) \circ (\mathbf{W}_{\text{FOG}}(\zeta) \cdot \mathbf{W}_{\text{FOG}}(\zeta)) \quad (6)$$

where $\circ$ denotes element-wise multiplication. We name $\mathcal{G}_{\text{FOG}}(\zeta)$ and $\mathcal{G}_{\text{SOG}}(\zeta)$ as the extended compatibility graphs, since their edge relations are parameterized by a scale variable ζ . As second-order graphs have been empirically validated for the 3D-3D registration [11], we adopt $\mathcal{G}_{\text{SOG}}(\zeta)$ in the following pruning method.

4.2 Pruning based on Extended Compatibility Graphs

Pruning prototype. Extended Sim(3) registration is an underdetermined problem and cannot be solved directly. To tackle this challenge, our pruning prototype aims to **approximate** the **extended Sim(3)** registration as a **standard SE(3)** registration by scale-correcting correspondences using a family of graphs $\{\mathcal{G}_{\text{SOG}}(\zeta)\}_{\zeta}$, as illustrated in Fig. 2.

Specifically, we first analyze the maximum clique (MAC) of the extended compatibility graph, denoted as $\text{Mac}(\mathcal{G}_{\text{SOG}}(\zeta))$. Under the inlier set maximization assumption [6, 37], $\text{Mac}(\mathcal{G}_{\text{SOG}}(\zeta))$ contains potential inliers. From the definition of extended compatibility graph in Eq. 5, if $\mathbf{d}_i = (\mathbf{q}_i, \mathbf{p}_i) \in \text{Mac}(\mathcal{G}_{\text{SOG}}(\zeta))$, $\mathbf{e}_i \triangleq (\zeta \cdot \mathbf{q}_i, \mathbf{p}_i)$ is an inlier candidate obeying the SE(3) transformation ($\mathbf{p}_i = \mathbf{R}(\zeta \mathbf{q}_i) + \mathbf{t}$). We write $\mathbf{e}_i = \varphi(\mathbf{d}_i, \zeta)$ and $\mathbf{e}_i \in \varphi(\text{Mac}(\mathcal{G}_{\text{SOG}}(\zeta)), \zeta)$ where $\varphi(\cdot)$ is called the scale-correction function. All such $\mathbf{e}_i$ yield a new correspondence set:

$$\{\mathbf{e}_j\}_{j=1}^M = \cup_{\zeta \in (0, +\infty)} \varphi(\text{Mac}(\mathcal{G}_{\text{SOG}}(\zeta)), \zeta) \quad (7)$$

Using function $\varphi(\cdot)$, the extended Sim(3) is converted as an SE(3) registration problem over $\{\mathbf{e}_j\}_{j=1}^M$. The set $\{\mathbf{e}_j\}_{j=1}^M$ has two advantages: (i) through maximum clique searching, inlier ratio of $\{\mathbf{e}_j\}_{j=1}^M$ is **higher than** $\{\mathbf{d}_i\}_{i=1}^N$; (ii) the inlier constrain in $\{\mathbf{e}_j\}_{j=1}^M$ is simplified as an SE(3) transformation, which is **easier** to solve than the extended Sim(3) transformation.

Once $\{\mathbf{e}_j\}_{j=1}^M$ are obtained, camera pose and inliers are robustly estimated:

$$\tilde{\mathbf{R}}, \tilde{\mathbf{t}} = \text{SE}(3)_{\text{-Reg}}(\{\mathbf{e}_j\}_{j=1}^M) \quad (8)$$

where $\tilde{\mathbf{R}}, \tilde{\mathbf{t}}$ are estimated camera pose. $\text{SE}(3)_{\text{-Reg}}(\cdot)$ stands for any off-the-shelf SE(3) registration algorithm. With $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{t}}$, inlier labels $\{w_i\}_{i=1}^N$ are identified by Eq. 1. Thus, the proposed pruning prototype (Eqs. 7 and 8) provides a feasible solution to Eq. 4.

Pruning algorithm. Built upon the above pruning prototype, we propose an efficient algorithm Ex-Sim(3)-Reg with three steps: (i) scale hypothesis sampling, (ii) inlier candidate searching, and (iii) pose estimation (Fig. 2). The **first step** is to sample ζ , since exhaustive search over $\zeta \in (0, +\infty)$ is infeasible. From Eq. 3, a coarse $\tilde{s}$ is estimated by the adaptive voting algorithm [30]. We heuristically restrict the search range to $\zeta \in (\tilde{s} - \Delta s, \tilde{s} + \Delta s)$ where Δs is a hyperparameter. We then uniformly sample K scales $\{\zeta_k\}_{k=1}^K$ from this interval and build the corresponding graphs $\{\mathcal{G}_{\text{SOG}}(\zeta_k)\}_{k=1}^K$ via Eqs. 5 and 6. The **second step** is to construct $\{\mathbf{e}_j\}_{j=1}^M$ from Eq. 7 where $\zeta \in (0, +\infty)$ is replaced by $\{\zeta_k\}_{k=1}^K$. The **third step** is to estimate pose and inliers. Camera pose is first estimated from Eq. 8 and then refined with RANSAC based P3P algorithm [24]:

$$\tilde{\mathbf{R}}_{\text{opt}}, \tilde{\mathbf{t}}_{\text{opt}}, \{w_i\}_{i=1}^N = \text{RANSAC_P3P}(\{\mathbf{c}_i\}_{i=1}^N, \tilde{\mathbf{R}}, \tilde{\mathbf{t}}) \quad (9)$$

In Eq. 9, given the reliable pose $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{t}}$ as priors, RANSAC based P3P [24] first pre-filters the outliers with the reprojection error larger than δ_{thr} and then estimate the more accurate pose $\tilde{\mathbf{R}}_{\text{opt}}, \tilde{\mathbf{t}}_{\text{opt}}$ using the remained inliers. The ablation study of δ_{thr} is provided in Sec. 5.3. More details of the algorithm are presented in the supplemental materials.

4.3 Theoretical Analysis of Pruning Prototype

To further justify the effectiveness of our method, we introduce the following key concept:

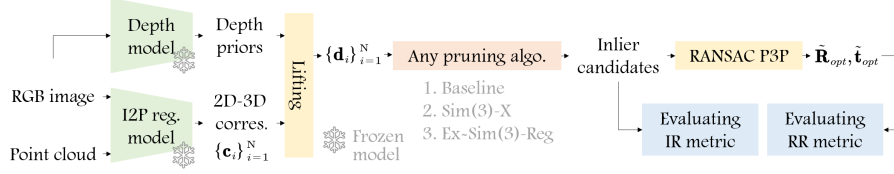


Fig. 3: Evaluation protocol. Inlier Ratio (IR) measures the quality of inlier candidates, while Registration Recall (RR) evaluates the accuracy of the estimated rotation $\hat{\mathbf{R}}_{opt}$ and translation $\hat{\mathbf{t}}_{opt}$. **Baseline** denotes no pruning. **Sim(3)-X** estimates a global scale $\tilde{s}$ [30], transforms correspondences from $\{\mathbf{d}_i\}_{i=1}^N$ to $\{(\tilde{s}\mathbf{q}_i, \mathbf{p}_i)\}_{i=1}^N$, and then applies an SE(3) solver $\mathbf{X}$ to identify inliers. A complete protocol is in the supplemental materials.

Ideal extended compatibility graph. If $\{s_i\}_{i=1}^N$ in Eq. 4 are known, we could construct the ideal extended compatibility graphs $\mathcal{G}_{\text{FOG}}(\{s_i\}_{i=1}^N)$, $\mathcal{G}_{\text{SOG}}(\{s_i\}_{i=1}^N)$. Edge matrix $\mathbf{W}_{\text{FOG}}(\{s_i\}_{i=1}^N)$ is defined analogously to Eq. 5 where $d_{ij}(\zeta)$ is replaced by $\rho_{ij}(s_i, s_j) = \|\mathbf{p}_i - \mathbf{p}_j\|_2 - \|s_i \cdot \mathbf{q}_i - s_j \cdot \mathbf{q}_j\|_2$. Similar to Eq. 6, $\mathbf{W}_{\text{SOG}}(\{s_i\}_{i=1}^N)$ is then computed as [11]:

$$\mathbf{W}_{\text{SOG}}(\{s_i\}_{i=1}^N) = \mathbf{W}_{\text{FOG}}(\{s_i\}_{i=1}^N) \circ (\mathbf{W}_{\text{FOG}}(\{s_i\}_{i=1}^N) \mathbf{W}_{\text{FOG}}(\{s_i\}_{i=1}^N)) \quad (10)$$

Under the inlier set maximization assumption [6, 37], solving the extended Sim(3) registration in Eq. 4 can be interpreted as searching for the maximum clique in the ideal extended compatibility graph $\mathcal{G}_{\text{SOG}}(\{s_i\}_{i=1}^N)$.

Relation between ideal and practical graphs. Based on the above, we investigate the relationship between the ideal graph $\mathcal{G}_{\text{SOG}}(\{s_i\}_{i=1}^N)$ and the practical graphs $\{\mathcal{G}_{\text{SOG}}(\zeta_k)\}_{k=1}^K$. Let $\mathcal{F}$ denote the ground truth inlier set. We define a subset $\mathcal{F}_k \subseteq \mathcal{F}$ that if $\mathbf{d}_i$ and $\mathbf{d}_j \in \mathcal{F}_k$, there exists a unique ζ_k satisfying $\rho_{ij}(\zeta_k) \leq d_{thr} < \rho_{ij}(\zeta_l)$ if $l \neq k$. Under the inlier set maximization assumption [6, 37], we have $\mathcal{F}_k \subseteq \text{Mac}(\mathcal{G}_{\text{SOG}}(\zeta_k))$ and $\mathcal{F} \subseteq \text{Mac}(\mathcal{G}_{\text{SOG}}(\{s_i\}_{i=1}^N))$. This leads to the following relation:

$$\cup_{k=1}^K \mathcal{F}_k \subseteq \text{Mac}(\mathcal{G}_{\text{SOG}}(\{s_i\}_{i=1}^N)) \cap (\cup_{k=1}^K \text{Mac}(\mathcal{G}_{\text{SOG}}(\zeta_k))) \quad (11)$$

This shows that the maximum cliques of ideal and practical graphs overlap on $\cup_{k=1}^K \mathcal{F}_k$. In short, the union of practical graphs $\{\mathcal{G}_{\text{SOG}}(\zeta_k)\}_{k=1}^K$ serves as an approximation of the ideal graph $\mathcal{G}_{\text{SOG}}(\{s_i\}_{i=1}^N)$, which theoretically validates Ex-Sim(3)-Reg. In addition, Eq. 11 reveals that our solution only searches for inliers within $\cup_{k=1}^K \mathcal{F}_k$, while the various experiments in Sec. 5.2 verify that this subset is sufficient to achieve robust pruning in most real-world scenarios.

5 Experiments and Discussions

5.1 Experimental Setup

To evaluate the accuracy of 2D-3D correspondence pruning, we conduct a series of experiments on the public indoor datasets, including 7-Scenes [16], RGBD-v2 [18], ScanNet [13], and TUM [25]. To prepare the pruning dataset, we adopt

Table 2: Correspondence pruning performance on the 7-Scenes dataset [16] under the disturbed condition. **Bold** indicates the best mean performance across scenes.

Method	SC-1	SC-2	SC-3	SC-4	SC-5	SC-6	Mean
<i>Inlier Ratio</i> $\uparrow$							
Baseline	0.295	0.257	0.325	0.265	0.227	0.228	0.266
Sim(3)-FastMAC	0.374	0.316	0.379	0.328	0.296	0.306	0.333
Sim(3)-SC2-PCR	0.481	0.435	0.431	0.446	0.436	0.475	0.451
Sim(3)-SC2-PCR++	0.295	0.254	0.325	0.259	0.229	0.251	0.268
Sim(3)-MAC	0.372	0.313	0.365	0.320	0.307	0.322	0.333
Sim(3)-TurboReg	0.369	0.311	0.373	0.316	0.306	0.323	0.333
Ex-Sim(3)-Reg	0.576	0.553	0.601	0.563	0.487	0.510	0.548 (+28.2%)
<i>Registration Recall</i> $\uparrow$							
Baseline	0.385	0.247	0.083	0.255	0.429	0.357	0.292
Sim(3)-FastMAC	0.246	0.192	0.251	0.170	0.411	0.429	0.283
Sim(3)-SC2-PCR	0.323	0.247	0.250	0.319	0.369	0.143	0.275
Sim(3)-SC2-PCR++	0.046	0.137	0.083	0.043	0.060	0.071	0.073
Sim(3)-MAC	0.338	0.260	0.167	0.234	0.440	0.214	0.275
Sim(3)-TurboReg	0.400	0.327	0.170	0.298	0.435	0.429	0.343
Ex-Sim(3)-Reg	0.415	0.493	0.500	0.383	0.542	0.500	0.472 (+18.0%)

a pre-trained I2P registration model MATR [20] to generate 2D-3D correspondences under three challenging settings (disturbed, unseen, and low-inlier conditions). More details of these settings are provided in the supplemental materials.

For fair comparison, our method is compared with the state-of-the-art non-learning correspondence pruning methods across four datasets, where the evaluation procedure is provided in Fig. 3. **Baseline** is the default pruning method. Pruning method with regular Sim(3) registration is marked as **Sim(3)-X** where **X** denotes the arbitrary SE(3) registration algorithm, like SC2-PCR [11], SC2-PCR++ [10], MAC [37], FastMAC [38], and TurboReg [29]. In the depth priors based methods, depths are predicted from RGB images using Depth Anything v2 (ViT-Large) [32]. Implementation details and hyperparameter configurations of the proposed method are provided in the supplemental materials. $\mathbf{R}_{gt}$ and $\mathbf{t}_{gt}$ are the ground truth camera pose. Two metrics are used for evaluation (computation details refer to the appendix in [20]):

- Inlier Ratio (IR) evaluates the ratio of inliers from inlier candidates as shown in Fig. 3. A 2D-3D correspondence $(\mathbf{l}_i, \mathbf{p}_i)$ can be regarded as inlier if the 3D distance between points $\mathbf{R}_{gt}\mathbf{p}_i + \mathbf{t}_{gt}$ and $d_{i,gt} \cdot \pi^{-1}(\mathbf{l}_i|\mathbf{K})$ ($d_{i,gt}$ is the ground truth depth of the pixel $\mathbf{l}_i$) is below a default threshold of 10cm.
- Registration Recall (RR) evaluates the accuracy of $\tilde{\mathbf{R}}_{opt}, \tilde{\mathbf{t}}_{opt}$. A successful registration is identified if the average 3D distance between 3D point clouds transformed by $(\tilde{\mathbf{R}}_{opt}, \tilde{\mathbf{t}}_{opt})$ and $(\mathbf{R}_{gt}, \mathbf{t}_{gt})$ is less than a default threshold of 5cm.

Table 3: Correspondence pruning performance on the TUM dataset [25] under the disturbed condition. **Bold** indicates the best mean performance.

Method	Freiburg_01	Freiburg_02	Freiburg_03	Mean
<i>Inlier Ratio</i> $\uparrow$				
Baseline	0.242	0.265	0.183	0.230
Sim(3)-FastMAC	0.317	0.363	0.241	0.307
Sim(3)-SC2-PCR	0.387	0.415	0.331	0.378
Sim(3)-SC2-PCR++	0.267	0.339	0.235	0.280
Sim(3)-MAC	0.323	0.355	0.231	0.303
Sim(3)-TurboReg	0.316	0.347	0.250	0.304
Ex-Sim(3)-Reg	0.428	0.548	0.480	0.485 (+25.5%)
<i>Registration Recall</i> $\uparrow$				
Baseline	0.160	0.040	0.152	0.117
Sim(3)-FastMAC	0.092	0.037	0.121	0.083
Sim(3)-SC2-PCR	0.067	0.037	0.128	0.077
Sim(3)-SC2-PCR++	0.050	0.029	0.058	0.046
Sim(3)-MAC	0.109	0.040	0.112	0.087
Sim(3)-TurboReg	0.118	0.049	0.125	0.097
Ex-Sim(3)-Reg	0.277	0.198	0.432	0.302 (+18.5%)

5.2 Main Comparisons

Performances under disturbed condition. We evaluate the compared methods in the stress-test scenarios. We simulate the disturbed condition and collect the dataset in two steps: (1) distorting color features of point cloud by the common adversarial attack algorithm [3]; (2) generating correspondences for I2P pruning by inferring MATR [20] with RGB images and distorted point clouds. Results on the 7-Scenes dataset [16] are provided in Table 2, where MATR [20] was pre-trained on the Chess scene and generated the pruning dataset on the other six scenes. Our method improves IR by **28.2%** and RR by **18.0%** compared to the baseline. Consistent improvements are observed on the TUM dataset [25] (Table 3), where our approach achieves substantial RR improvements (**20%**) on Freiburg_02, while all baselines remain below 5%. This demonstrates strong robustness against feature corruption. Results with the different depth prediction methods are provided in the supplemental materials.

Performances under unseen condition. We further assess cross-domain generalization by training on one dataset and testing on unseen target datasets. In the 7-Scenes $\rightarrow$ RGBD-v2 and 7-Scenes $\rightarrow$ ScanNet experiments, our method achieves RR gains of **13.1%** and **20.4%**, respectively (Tables 4 and 5). The larger improvement on ScanNet reflects its greater scene diversity and varying depth-scale distortions, which pose greater challenges to existing methods. Standard Sim (3)-based pruning methods fail to outperform the baseline, because they overlook the noises in predicted depths, overfitting to false-positive inliers. In contrast, our extended Sim(3) based method explicitly compensates

Table 4: Cross-dataset generalization from 7-Scenes [16] to RGBD-v2 [18]. **Bold** indicates the best mean performance. Results demonstrate robustness to domain shift in depth priors and correspondence distributions.

Method	RG-1	RG-2	RG-3	RG-4	RG-5	RG-6	RG-7	Mean
<i>Inlier Ratio</i> $\uparrow$								
Baseline	0.242	0.248	0.223	0.224	0.273	0.282	0.259	0.250
Sim(3)-FastMAC	0.309	0.326	0.275	0.283	0.354	0.341	0.350	0.320
Sim(3)-SC2-PCR	0.413	0.415	0.352	0.365	0.367	0.369	0.373	0.379
Sim(3)-SC2-PCR++	0.205	0.252	0.239	0.231	0.309	0.322	0.248	0.258
Sim(3)-MAC	0.322	0.317	0.281	0.281	0.307	0.319	0.308	0.305
Sim(3)-TurboReg	0.391	0.315	0.280	0.279	0.307	0.321	0.306	0.314
Ex-Sim(3)-Reg	0.466	0.455	0.452	0.428	0.471	0.472	0.442	0.455 (+20.5%)
<i>Registration Recall</i> $\uparrow$								
Baseline	0.237	0.240	0.226	0.444	0.267	0.159	0.188	0.252
Sim(3)-FastMAC	0.263	0.280	0.226	0.407	0.173	0.159	0.167	0.239
Sim(3)-SC2-PCR	0.263	0.320	0.355	0.333	0.280	0.058	0.188	0.257
Sim(3)-SC2-PCR++	0.026	0.031	0.055	0.061	0.253	0.101	0.229	0.108
Sim(3)-MAC	0.342	0.201	0.323	0.370	0.200	0.043	0.104	0.226
Sim(3)-TurboReg	0.289	0.240	0.324	0.369	0.160	0.232	0.188	0.257
Ex-Sim(3)-Reg	0.500	0.320	0.484	0.556	0.280	0.333	0.208	0.383 (+13.1%)

Table 5: Cross-dataset evaluation from 7-Scenes [16] to ScanNet [13]. ‘N/A’ indicates failure to return a result within 1 second, reflecting computational instability under severe outlier contamination.

Method	SN-1	SN-2	SN-3	SN-4	SN-5	SN-6	SN-7	Mean
<i>Inlier Ratio</i> $\uparrow$								
Baseline	0.463	0.469	0.388	0.269	0.474	0.408	0.390	0.409
Sim(3)-FastMAC	0.553	0.570	0.478	0.344	0.566	0.501	0.566	0.511
Sim(3)-SC2-PCR	0.625	0.616	0.552	0.398	0.629	0.594	0.644	0.579
Sim(3)-SC2-PCR++	0.473	0.487	0.377	0.297	0.463	0.442	0.472	0.430
Sim(3)-MAC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Sim(3)-TurboReg	0.517	0.506	0.434	0.338	0.521	0.482	0.520	0.474
Ex-Sim(3)-Reg	0.694	0.747	0.674	0.423	0.743	0.639	0.715	0.662 (+25.3%)
<i>Registration Recall</i> $\uparrow$								
Baseline	0.611	0.444	0.301	0.205	0.502	0.455	0.499	0.431
Sim(3)-FastMAC	0.444	0.445	0.400	0.0	0.577	0.091	0.501	0.351
Sim(3)-SC2-PCR	0.446	0.222	0.195	0.103	0.385	0.088	0.378	0.260
Sim(3)-SC2-PCR++	0.172	0.155	0.0	0.099	0.308	0.0	0.125	0.123
MAC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Sim(3)-TurboReg	0.278	0.281	0.304	0.101	0.462	0.455	0.250	0.304
Ex-Sim(3)-Reg	0.833	0.722	0.600	0.186	0.808	0.545	0.750	0.635 (+20.4%)

for noisy depth priors, yielding stable I2P registration performances. Additional metrics of the pose errors are reported in the supplemental materials.

Table 6: Correspondence pruning performance on the RGBD-v2 dataset [18] under low-inlier conditions. We restrict the inlier number to simulate extreme outlier scenarios. These highlight robustness when geometric constraints are limited.

Method	RG-11	RG-12	RG-13	RG-14	Mean
<i>Inlier Ratio</i> $\uparrow$					
Baseline	0.195	0.271	0.075	0.123	0.166
Sim(3)-FastMAC	0.241	0.318	0.085	0.138	0.195
Sim(3)-SC2-PCR	0.334	0.392	0.112	0.159	0.249
Sim(3)-SC2-PCR++	0.231	0.294	0.085	0.130	0.185
Sim(3)-MAC	0.244	0.311	0.093	0.135	0.196
Sim(3)-TurboReg	0.239	0.309	0.091	0.134	0.193
Ex-Sim(3)-Reg	0.444	0.514	0.144	0.206	0.327 (+16.1%)
<i>Registration Recall</i> $\uparrow$					
Baseline	0.347	0.443	0.077	0.314	0.295
Sim(3)-FastMAC	0.431	0.533	0.128	0.236	0.332
Sim(3)-SC2-PCR	0.556	0.484	0.171	0.210	0.355
Sim(3)-SC2-PCR++	0.153	0.410	0.111	0.144	0.204
Sim(3)-MAC	0.292	0.607	0.145	0.295	0.335
Sim(3)-TurboReg	0.486	0.484	0.103	0.303	0.344
Ex-Sim(3)-Reg	0.708	0.852	0.231	0.376	0.542 (+24.7%)

Table 7: Cross-domain I2P registration performance on the ScanNet dataset [13]. We integrate the proposed Ex-Sim(3)-Reg (Ours) into different baselines, which demonstrates the superior improvements.

Method	Matr	+Ours	FreeReg	+Ours	Top-I2P	+Ours	MinCD	+Ours	Diff-I2P	+Ours
IR	0.451	0.662	0.568	0.690	0.493	0.674	0.464	0.678	0.521	0.682
RR	0.842	0.943	0.780	0.932	0.901	0.950	0.898	0.957	0.914	0.963

Performances under low-inlier condition. We evaluate the pruning performance under the condition of extremely few inliers. A low-inlier correspondence dataset is generated from the RGBD-v2 dataset [18], containing sparse point clouds and low-quality RGB images. This setting is created by inferring I2P correspondences from a model trained with insufficient supervision. In such a condition, the traditional RANSAC-based P3P and regular Sim(3)-based pruning methods typically fail due to insufficient geometric constraints. As presented in Table 6, our method achieves the highest IR and RR, exceeding 50% RR even under severe correspondence sparsity. This is because Ex-Sim(3)-Reg leverages a family of extended compatibility graphs to reliably discover potential inliers, thus enabling robust registration.

Qualitative comparison. Figure 4 visualizes correspondence pruning results. With the extended Sim(3) registration model, our method recovers denser and geometrically consistent inlier sets.

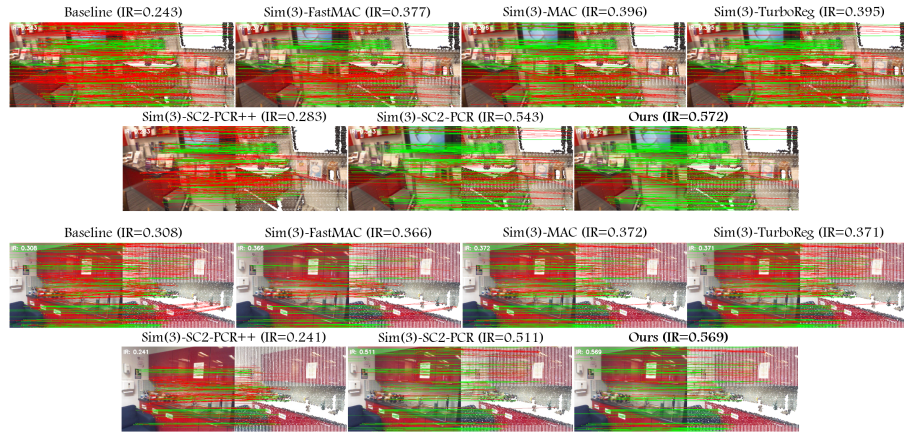


Fig. 4: Qualitative comparison on challenging scenes. Competing methods either remove excessive inliers or retain geometrically inconsistent matches. Our method preserves denser and spatially coherent inlier sets.

Table 8: Effect of the number of sub-intervals K on IR, RR, and runtime. Increasing K improves accuracy but with increased computational overhead.

K	4	6	8	10	12
IR/RR	0.505/0.362	0.542/0.394	0.548/ 0.472	0.573/0.448	0.682 /0.427
Time	57.6ms	65.4ms	73.6ms	81.7ms	89.5ms

Comparison to I2P registration. Moreover, we integrate the proposed pruning method into the state-of-the-art I2P registration methods, such as FreeReg [28], Top-I2P [1], and MinCD [2]. Methods except FreeReg (a zero-shot method) are all trained on 7-Scenes [16] but tested on ScanNet [13]. From Table 7, it is observed that Ex-Sim(3)-Reg improves the performance of current methods, which validates the practical value of our pruning pipeline.

Results analysis. Several patterns can be observed from the above results. Standard Sim(3)-based pruning ignores depth prediction errors, leading to only marginal or even negative RR gains. Sim(3)-SC2-PCR++ [10] uses a strict inlier criterion designed for 3D registration, but fails to account for point cloud distortion from noisy depth maps, resulting in unstable I2P registration. Sim(3)-TurboReg [29] focuses on identifying the optimal cliques with minimal elements, and thus exhibits robustness to the noisy depths than others based on regular Sim(3) registration. Different from the previous methods, the proposed extended Sim(3) based pruning method naturally models depth noise, explicitly corrects noisy correspondences, thus achieving the most consistent and robust performance across all settings.

Table 9: Impact of pixel reprojection threshold δ_{thr} on IR, RR, and relative translation error (RTE). Larger thresholds improve robustness but may admit additional outliers, revealing the robustness–precision trade-off.

δ_{thr}	8	12	16	20	24
IR/RR	0.513/0.380	0.535/0.466	0.548/0.472	0.521/0.384	0.505/0.372
RTE	0.072m	0.068m	0.067m	0.070m	0.071m

Table 10: Ablation studies on MAC search and scale-correction function. Removing the MAC search significantly degrades inlier consistency, while omitting the correction function reduces robustness to depth noise. Runtime comparison illustrates the computational trade-offs.

MAC [37]	FastMAC [38]	Scale-correction function $\varphi(\cdot)$	RR	Average Runtime/ms
	✓		0.311	72.4
✓		✓	0.485	204.8
	✓	✓	0.472	73.6

5.3 Ablation Studies

We investigate the effect of key hyperparameters and components of the proposed method on the noisy 7-Scenes dataset [16].

Number of sub-intervals K . Increasing K can enlarge the upper bound of inliers, since the size of $\cup_{k=1}^K \mathcal{F}_k$ is growing with K . As shown in Table 8, performance saturates when $K > 6$, while the computational cost continues to grow. We thus adopt $K = 6$ to balance accuracy and efficiency.

Pixel error threshold δ_{thr} . This parameter can weigh up the robustness and accuracy of Ex-Sim(3)-Reg. A smaller δ_{thr} benefits the precise poses but degrades robustness when the initial pose $\hat{\mathbf{R}}, \hat{\mathbf{t}}$ are noisy. Table 9 shows that $\delta_{thr} = 16$ provides the best trade-off between IR, RR, and RTE metrics, and is therefore used in our experiments.

MAC search and scale-correction. We further investigate the effects of using FastMAC [38] versus MAC [37], and scale-correction function. As summarized in Table 10, MAC achieves slightly higher RR but is $2.78\times$ slower. FastMAC offers a favorable accuracy-efficiency trade-off. Correction function substantially improves accuracy across both variants, confirming its necessity for reliable correspondence pruning.

6 Conclusions

In this paper, we reformulate depth prior-based 2D-3D correspondence pruning as an extended Sim(3) registration problem. Different from regular Sim(3), the extended Sim(3) can account for errors in depth priors, thus enabling the robust pruning performance. However, extended Sim(3) registration is an under-determined problem, and no previous work has specifically studied it. To address

this problem, we propose a heuristic yet effective pruning algorithm named Ex-Sim(3)-Reg. By detecting and correcting maximum cliques from a family of extended compatibility graphs, Ex-Sim(3)-Reg simplifies extended Sim(3) to the standard SE(3) registration with a high-inlier set, allowing accurate and stable camera pose estimation. Moreover, theoretical analysis is conducted to justify the effectiveness of Ex-Sim(3)-Reg. Finally, experiments on four public datasets verify that our approach significantly outperforms state-of-the-art methods.

Limitation and future work. Despite the superior performances, our method can recover at most the inlier set $\cup_{k=1}^K \mathcal{F}_k$. Even when K is $+\infty$, $\cup_{k=1}^{+\infty} \mathcal{F}_k$ remains a subset of the true inlier set. To search for more potential inliers, we will refine the construction and usage of extended compatibility graphs in future work.

Acknowledgments. The authors greatly thank Prof. You Yang for his suggestions. This work is partially supported by the National Natural Science Foundation of China (Grand ID: 62502171, 62372377) and China Postdoctoral Science Foundation (Grand ID: 2024M761014, GZC20252285)

References

1. An, P., Yang, J., Peng, M., et al.: Top-I2P: Explore open-domain image-to-point cloud registration using topology relationship. In: Proceedings of International Joint Conference on Artificial Intelligence. pp. 1–9 (2025)
2. An, P., Yang, J., Peng, M., Yang, Y., Liu, Q., Wu, X., Nan, L.: Mincd-pnp: Learning 2d-3d correspondences with approximate blind pnp. In: Proceedings of International Conference on Computer Vision. pp. 1–11 (2025)
3. Bian, Y., Tian, S., Liu, X.: iba: Backdoor attack on 3d point cloud via reconstructing itself. *IEEE Trans. Inf. Forensics Secur.* **19**, 7994–8008 (2024)
4. Bie, L., Pan, S., Li, S., et al.: Graphi2p: Image-to-point cloud registration with exploring pattern of correspondence via graph learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 22161–22171 (2025)
5. Campbell, D., Liu, L., Gould, S.: Solving the blind perspective-n-point problem end-to-end with robust differentiable geometric optimization. In: Proceedings of European Conference on Computer Vision. pp. 244–261 (2020)
6. Campbell, D., Petersson, L., Kneip, L., Li, H.: Globally-optimal inlier set maximisation for camera pose and correspondence estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(2), 328–342 (2020)
7. Cattaneo, D., Vaghi, M., Ballardini, A.L., Fontana, S., Sorrenti, D.G., Burgard, W.: Cmrnet: Camera to lidar-map registration. In: Proceedings of IEEE Intelligent Transportation Systems Conference. pp. 1283–1289 (2019)
8. Cattaneo, D., Valada, A.: Cmrnext: Camera to lidar matching in the wild for localization and extrinsic calibration. *IEEE Trans. Robotics* **41**, 1995–2013 (2025)
9. Chen, H., Wang, P., Wang, F., et al.: Epro-pnp: Generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2771–2780 (2022)
10. Chen, Z., Sun, K., Yang, F., Guo, L., Tao, W.: Sc²-pcr++: Rethinking the generation and selection for efficient and robust point cloud registration. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(10), 12358–12376 (2023)

11. Chen, Z., Sun, K., Yang, F., Tao, W.: Sc²-pcr: A second order spatial compatibility for efficient and robust point cloud registration. In: Proceedings of Conference on Computer Vision and Pattern Recognition. pp. 13211–13221 (2022)
12. Cheng, Z., Deng, J., Li, X., et al.: Bridge 2d-3d: Uncertainty-aware hierarchical registration network with domain alignment. In: Proceedings of AAAI Conference on Artificial Intelligence. pp. 2491–2499 (2025)
13. Dai, A., Chang, A.X., Savva, M., et al.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 2432–2443 (2017)
14. David, P., DeMenthon, D., Duraiswami, R., Samet, H.: Softposit: Simultaneous pose and correspondence determination. *Int. J. Comput. Vis.* **59**(3), 259–284 (2004)
15. Feng, M., Hu, S., Ang, M.H., Lee, G.H.: 2D3D-Matchnet: Learning to match keypoints across 2d image and 3d point cloud. In: Proceedings of IEEE International Conference on Robotics and Automation. pp. 4790–4796 (2019)
16. Glocker, B., Izadi, S., Shotton, J., Criminisi, A.: Real-time RGB-D camera relocalization. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality. pp. 173–179 (2013)
17. Kang, S., Liao, Y., Li, J., et al.: Cofii2p: Coarse-to-fine correspondences-based image to point cloud registration. *IEEE Robotics Autom. Lett.* **9**(11), 10264–10271 (2024)
18. Lai, K., Bo, L., Fox, D.: Unsupervised feature learning for 3d scene labeling. In: Proceedings of IEEE International Conference on Robotics and Automation. pp. 3050–3057 (2014)
19. Leroy, V., Cabon, Y., Revaud, J.: Grounding image matching in 3d with mast3r. In: Proceedings of European Conference on Computer Vision. pp. 71–91 (2024)
20. Li, M., Qin, Z., Gao, Z., et al.: 2D3D-MATR: 2D-3D matching transformer for detection-free registration between images and point clouds. In: Proceedings of IEEE Conference on Computer Vision. pp. 1–10 (2023)
21. Li, X., Yang, W., Deng, J., et al.: Implicit correspondence learning for image-to-point cloud registration. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 16922–16931 (2025)
22. Liang, Y., Hu, Y., Shao, W., Fu, Y.: Distilling monocular foundation model for fine-grained depth completion. In: Proceedings of International Conference on Computer Vision and Pattern Recognition. pp. 22254–22265 (2025)
23. Murai, R., Dexheimer, E., Davison, A.J.: Mast3r-slam: Real-time dense slam with 3d reconstruction priors. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1–10 (2025)
24. Persson, M., Nordberg, K.: Lambda twist: An accurate fast robust perspective three point (P3P) solver. In: Proceedings of European Conference on Computer Vision. vol. 11208, pp. 334–349 (2018)
25. Sturm, J., Engelhard, N., Endres, F., et al.: A benchmark for the evaluation of RGB-D SLAM systems. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 573–580 (2012)
26. Sun, Y., Cheng, C., Zhang, Y., et al.: Circle loss: A unified perspective of pair similarity optimization. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6397–6406 (2020)
27. Wang, B., Chen, C., Cui, Z., et al.: P2-Net: Joint description and detection of local features for pixel and point matching. In: Proceedings of IEEE International Conference on Computer Vision. pp. 15984–15993 (2021)

28. Wang, H., Liu, Y., Wang, B., et al.: Freereg: Image-to-point cloud registration leveraging pretrained diffusion models and monocular depth estimators. In: Proceedings of International Conference on Learning Representation. pp. 1–24 (2024)
29. Yan, S., Shi, P., Zhao, Z., Wang, K., Cao, K., Wu, J., Li, J.: Turboreg: Turboclique for robust and efficient point cloud registration. In: Proceedings of International Conference on Computer Vision. pp. 1–11 (2025)
30. Yang, H., Shi, J., Carlone, L.: TEASER: fast and certifiable point cloud registration. *IEEE Trans. Robotics* **37**(2), 314–333 (2021)
31. Yang, L., Kang, B., Huang, Z., et al.: Depth anything: Unleashing the power of large-scale unlabeled data. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10371–10381 (2024)
32. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything V2. In: Proceedings of Advances in Neural Information Processing Systems. pp. 1–30 (2024)
33. Yao, G., Xuan, Y., Li, X., Pan, Y.: Cmr-agent: Learning a cross-modal agent for iterative image-to-point cloud registration. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 13458–13465 (2024)
34. Ye, Y., Teng, X., Yang, H., Chen, S., Sun, Y., Bian, Y., Tan, T., Li, Z., Yu, Q.: 3mos: a multi-source, multi-resolution, and multi-scene optical-sar dataset with insights for multi-modal image matching. *Vis. Intell.* **3**(1) (2025)
35. Yu, H., Chen, K., Yang, W., Scherer, S.A., Xia, G.: I2d-loc++: Camera pose tracking in lidar maps with multi-view motion flows. *IEEE Robotics Autom. Lett.* **9**(9), 8162–8169 (2024)
36. Yu, Y., Liu, S., Pautrat, R., Pollefeys, M., Larsson, V.: Relative pose estimation through affine corrections of monocular depth priors. In: Proceedings of International Conference on Computer Vision and Pattern Recognition. pp. 16706–16716 (2025)
37. Zhang, X., Yang, J., Zhang, S., Zhang, Y.: 3d registration with maximal cliques. In: Proceedings of Conference on Computer Vision and Pattern Recognition. pp. 17745–17754. *IEEE* (2023)
38. Zhang, Y., Zhao, H., Li, H., Chen, S.: Fastmac: Stochastic spectral sampling of correspondence graph. In: Proceedings of Conference on Computer Vision and Pattern Recognition. pp. 17857–17867 (2024)
39. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000)
40. Zhou, Q., Agostinho, S., Osep, A., Leal-Taixé, L.: Is geometry enough for matching in visual localization? In: Proceedings of European Conference on Computer Vision. vol. 13670, pp. 407–425 (2022)