Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DSG-RANSAC: Local Depth-Support-Guided Random Sample Consensus for RGB-D Registration

Fujia Liu

fujia.liu@foxmail.com

University of Science and Technology Beijing

Zhihao Zhao

University of Science and Technology Beijing

Jianhong Yang

University of Science and Technology Beijing

Chaoyang Duan

University of Science and Technology Beijing

Jiahua Sun

University of Science and Technology Beijing

Research Article

Keywords: RGB-D point cloud registration, random sample consensus, local depth support, guided sampling, pose estimation

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11634708/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DSG-RANSAC: Local Depth-Support-Guided Random Sample Consensus for RGB-D Registration

Fujia Liu[1,2*], Zhihao Zhao[2], Jianhong Yang[1,2*], Chaoyang Duan[2], Jiahua Sun[2]

[1]Technical Support Center for Prevention and Control of Disastrous Accidents in Metal Smelting, University of Science and Technology Beijing, Beijing, 100083, China.
[2]School of Mechanical Engineering, University of Science and Technology Beijing, Beijing, 100083, China.

*Corresponding author(s). E-mail(s): fujia.liu@foxmail.com; yangjianhong@me.ustb.edu.cn;
Contributing authors: ustbzzh@163.com; D202410324@xs.ustb.edu.cn; 17392765121@163.com;

Abstract

Incorrect matches and unreliable depth observations affect pose estimation in RGB-D coarse registration. We propose depth-support-guided random sample consensus (DSG-RANSAC) to address this problem. The method constructs correspondence-level local depth support from valid-depth coverage and normalized depth differences in the neighborhoods of matched endpoints. This support guides three-correspondence sampling without replacement. Three-edge geometric prechecking, rigid pose estimation, truncated linear scoring, and inlier-based refinement then produce the registration transform. Experiments on TUM indoor scenes and ROBI industrial parts show a statistical association between depth support and reference geometric inliers. Guided sampling increases the proportions of all-inlier triplets and pose hypotheses that meet the reference accuracy threshold. Under equal time budgets, the method achieves lower rotation and translation errors on six TUM sequences and five ROBI part categories. When depth support is used to rank correspondences for progressive sample consensus (PROSAC), the all-inlier triplet rate increases from 71.9% to 81.5% on TUM and from 76.2% to 82.6% on ROBI. The qualified hypothesis rate increases from 53.6% to 69.6% and from 48.4% to 57.5%, respectively. These results show that local depth support provides a correspondence quality prior for different sampling frameworks.

Keywords: RGB-D point cloud registration, random sample consensus, local depth support, guided sampling, pose estimation

1 Introduction

Point cloud registration estimates spatial transformations between views to express local observations in a common coordinate system. It is an important component of 3D reconstruction and pose estimation [1]. In industrial vision, multi-view geometric processing of parts requires an initial pose from coarse registration. Tasks such as workpiece and blade measurement require registration under noise and missing observations [2, 3]. For RGB-D data, pixel matches can first be established between color images and then back-projected into 3D correspondences using depth and camera intrinsics. These correspondences are used to estimate the rigid transformation from the source view to the target view. Viewpoint

changes, occlusion, and repeated textures introduce incorrect matches, while missing depth and measurement noise affect the positions of the 3D endpoints. Estimating a pose from correspondences with unreliable observations is therefore a central problem in RGB-D coarse registration.

Random sample consensus (RANSAC) handles estimation with outliers by repeatedly drawing minimal sets, generating model hypotheses, and evaluating their support [4]. In 3D rigid registration, three noncollinear correspondences form a minimal set for pose estimation. Uniform sampling gives every correspondence the same selection probability. When correct correspondences are sparse, more attempts are needed to obtain a usable triplet [5]. Under a limited iteration or time budget, candidate generation and geometric feasibility thus affect the pose search.

Previous methods organize candidate search using different sources of information. PROSAC progressively expands the sampling pool according to a given correspondence quality ranking [6]. SLHG constructs local sampling graphs from correspondence compatibility and combines local and global information to generate hypotheses [5]. SC^2-PCR and GCC use second-order spatial compatibility and RGB-D geometric constraints, respectively, to select or organize correspondences [7, 8]. Both individual correspondence quality and geometric relationships between correspondences can therefore contribute to robust estimation. For 3D correspondences constructed from RGB-D pixel matches, matching quality primarily describes the association between image endpoints, whereas depth observations determine their back-projected spatial positions. This work investigates how depth observations around matched endpoints can provide correspondence-level quality information for pose search under a limited budget.

Depth measurement errors affect the positions of back-projected points. Previous work has examined the relationship between Kinect depth accuracy and measurement range [9]. Within an individual frame, neighborhoods of matched endpoints may also contain missing depth or abrupt depth changes. Local depth support describes these neighborhood observations. Subsequent geometric checks and pose evaluation determine whether multiple correspondences can jointly support a rigid motion.

We propose Depth-Support-Guided RANSAC (DSG-RANSAC). The method combines valid-depth coverage and normalized depth differences to compute local support for the source and target endpoints. The smaller endpoint support becomes the correspondence-level weight. This weight guides sampling of three correspondences without replacement. Candidate triplets undergo a three-edge geometric precheck, followed by pose estimation, truncated linear scoring, and inlier-based refinement. The main contributions are:

1. A correspondence-level local depth support measure that normalizes depth differences by the lateral scale of the back-projected neighborhood and combines valid coverage with depth consistency as a sampling prior.

2. A DSG-RANSAC estimation procedure that combines depth-support-guided sampling with edge-length consistency prechecking of triplets, with a precheck condition derived from the 3D residual threshold.

Experiments on TUM and ROBI examine the relationship between depth support and sampling quality, and evaluate registration error, computational budgets, different frontends, and comparisons with existing estimators. We also use depth support as a ranking prior for PROSAC and compare it with frontend quality to examine its use across sampling frameworks. Sect. 2 reviews related work. Sect. 3 presents the method, Sect. 4 reports the experiments, Sect. 5 discusses the results, and Sect. 6 concludes the paper.

2 Related work

2.1 Correspondence-driven RGB-D rigid registration

Correspondence-based rigid registration first establishes candidate associations between source and target observations, then estimates rotation and translation from these associations [1]. RGB-D images provide both appearance and depth. A frontend establishes pixel correspondences in the image domain, and a backend back-projects them into 3D point pairs using depth. Match correctness and endpoint depth quality jointly determine the resulting 3D constraints.

Conventional point cloud registration often obtains a coarse alignment and refines it with ICP or its variants [10–12]. Other work improves the

efficiency of correspondence search in ICP [13]. To handle partial overlap and outliers, researchers have developed registration methods based on quantile assignment, learned features, and cross-scene frameworks [1, 14–16].

Image correspondences can be generated by different feature detection and matching methods. SIFT uses scale-space keypoints and local gradient descriptors [17]. SuperPoint learns interest points and local descriptors [18], while Efficient LoFTR generates semi-dense matches using aggregated attention and correlation computation [19]. Learned matching methods also include SuperGlue for local feature association, detector-free LoFTR, and LightGlue for local feature matching [20–22]. Surface curvature has also been used to guide image local feature matching [23]. These frontends produce different correspondence counts and qualities, providing inputs for evaluating the same backend under different conditions.

Geometric features such as 3DMatch, FCGF, and D3Feat can also establish 3D correspondences [24–26]. PREDATOR focuses on overlapping regions, while GeoTransformer and REGTR use contextual and geometric information to establish correspondences [27–29]. For RGB-D data, colored point cloud registration, PointMBF, RGBD-Glue, and ViGG explore the combination of appearance and geometry [30–33].

Correspondence construction and registration estimation can also be designed jointly. Oğuz et al. formulate correspondence search as a quantile assignment problem and combine descriptors, transformation estimation, and local refinement in a coarse-to-fine pipeline [1]. This work studies robust pose estimation from 3D correspondences generated by image matching.

2.2 Quality-guided sampling and robust estimation

Guided sampling uses candidate quality to change the hypothesis search order or sampling probabilities. PROSAC starts with highly ranked correspondences and progressively expands the candidate pool [6]. Because its search depends on the input ranking, frontend quality and other available priors can both provide ranking information.

Robust estimation can also incorporate local optimization and model scoring. USAC organizes multiple sampling and verification strategies. Graph-Cut RANSAC optimizes inlier sets using spatial coherence. MAGSAC and MAGSAC++ construct model evaluation and optimization procedures by marginalizing over the noise scale [34–37].

Sampling priors can also be learned or derived from relationships between correspondences. NG-RANSAC learns a sampling distribution with model estimation included in its training objective [38]. SLHG first forms subsets based on correspondence compatibility, guides sampling within local compatibility graphs, and combines global information to generate pose hypotheses [5]. CG-SAC guides sampling using correspondence compatibility, while SAC-COT organizes triplet candidates as compatible triangles [39, 40].

Our prior is derived from the depth neighborhoods at both endpoints of an individual correspondence. Local support is computed separately for the source and target endpoints and aggregated into a correspondence-level weight. DSG-RANSAC uses these weights to sample triplets without replacement. We also use depth support to rank correspondences for PROSAC and compare it with frontend quality ranking under the same sampling schedule.

2.3 Depth observation reliability and geometric consistency checks

Depth observations directly determine the spatial positions of back-projected points. Belter et al. [41] analyze the effects of RGB-D depth measurements and image processing on the spatial uncertainty of point features. Khoshelham and Oude Elberink examine the accuracy and resolution of Kinect depth data and the variation of measurement error with distance [9]. In two-view matching, valid-depth coverage and local depth variation around an endpoint can be computed from the current depth map to describe its local observational support.

Geometric consistency uses relationships among multiple correspondences to determine whether they can jointly explain a rigid motion. Deep Global Registration learns correspondence confidence and combines it with weighted pose estimation, while PointDSC uses spatial consistency to filter correspondences [42, 43].

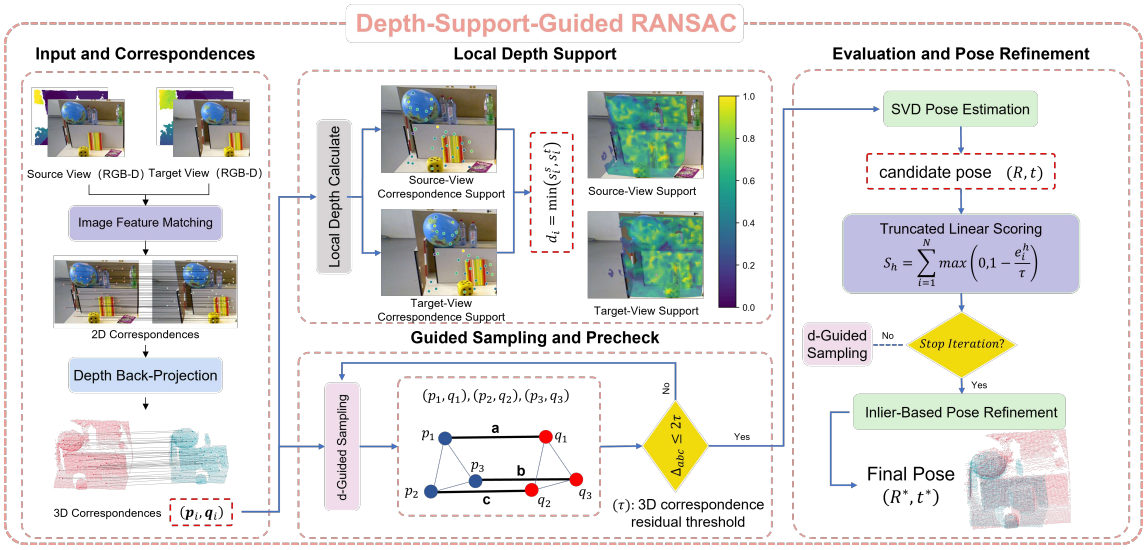


Fig. 1: Overview of Depth-Support-Guided RANSAC (DSG-RANSAC)

SC^2-PCR constructs second-order spatial compatibility from shared compatibility relationships and uses it for seed selection and consensus set construction [7]. GCC expresses 3D distance preservation between RGB-D correspondences as image-domain constraints for candidate verification [8].

We use local depth support and inter-correspondence geometric constraints at different stages. Local support determines selection weights before sampling. Geometric constraints check the edge-length consistency of each sampled minimal set. Candidates that pass the precheck and degeneracy checks generate pose hypotheses, which are evaluated using residuals over all correspondences.

3 Robust registration guided by local depth support

3.1 Problem formulation and framework

We consider two-view RGB-D rigid pose estimation with incorrect correspondences. Let D_s, D_t be the source and target depth maps and K_s, K_t their camera intrinsic matrices. The depth maps are registered to the images used for matching. The frontend provides the pixel correspondence set

$$\mathcal{M} = \{(\mathbf{u}_i^s, \mathbf{u}_i^t)\}_{i=1}^{N_0}. \tag{1}$$

For a pixel $\mathbf{u}$, a valid depth z is read from the depth map. Using the pixel coordinates $\mathbf{u}^{\mathrm{bp}} = (u^{\mathrm{bp}}, v^{\mathrm{bp}})^\top$ adopted for back-projection, the pinhole camera model gives

$$\mathbf{x} = zK^{-1}\widetilde{\mathbf{u}}^{\mathrm{bp}}, \qquad \widetilde{\mathbf{u}}^{\mathrm{bp}} = (u^{\mathrm{bp}}, v^{\mathrm{bp}}, 1)^\top. \tag{2}$$

Here, z is optical-axis depth, and the pixel coordinates and camera intrinsics use the same coordinate convention. Only correspondences with valid depth at both endpoints are retained. After unifying depth units, the 3D correspondence set is

$$\mathcal{C} = \{c_i = (\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^{N}, \tag{3}$$

The points $\mathbf{p}_i, \mathbf{q}_i$ are expressed in the source and target camera frames, respectively. The transformation $(R, \mathbf{t})$ maps source points to the target frame. Correct correspondences satisfy

$$\mathbf{q}_i = R\mathbf{p}_i + \mathbf{t} + \boldsymbol{\epsilon}_i, \qquad R \in \mathrm{SO}(3), \quad \mathbf{t} \in \mathbb{R}^3, \tag{4}$$

where $\boldsymbol{\epsilon}_i$ is the observation residual. For any candidate transformation, the geometric residual is

$$e_i(R, \mathbf{t}) = \|R\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2. \tag{5}$$

Fig. 1 summarizes DSG-RANSAC. Given two RGB-D frames and camera intrinsics, the method constructs correspondences, computes local depth support, performs guided sampling and geometric prechecking, and evaluates hypotheses before refining the pose.

Pixel correspondences are first established between the source and target RGB images. Matches with valid depth at both endpoints are retained and back-projected using the camera intrinsics to form $\mathcal{C}$. For each correspondence, support is computed from the source and target depth neighborhoods. The smaller endpoint support becomes its weight d_i. The support captures both valid-depth coverage and depth consistency; Sect. 3.2 describes its construction.

The algorithm then samples three distinct correspondences without replacement according to d_i. The three corresponding edges undergo a length-consistency precheck. Rejected candidates are discarded, and retained candidates undergo a degeneracy check. Every draw counts as an attempt. Triplets that pass these checks generate source-to-target pose hypotheses using SVD. Sect. 3.3 defines the sampling distribution and precheck condition.

For each valid pose hypothesis, residuals are evaluated over all 3D correspondences and scored with a truncated linear function. The highest-scoring hypothesis satisfying the inlier-count requirement is retained. Once the stopping condition is reached, a single SVD refinement on its inlier set produces the final rotation and translation. Estimation fails if no usable hypothesis is found or refinement fails. Sect. 3.4 details hypothesis evaluation and refinement.

3.2 Construction of local depth support

Endpoint support combines two observations: the coverage of valid depth in the neighborhood and the consistency of valid depth with the central observation. To place depth differences on a common scale, we first establish the lateral 3D scale of the pixel window and then compute neighborhood contributions from normalized depth differences.

We define support for a single endpoint and omit the view superscript. A square window of radius r is centered at the nearest integer pixel to endpoint $(\mathbf{u}_i, z_i)$. Removing the central integer pixel and positions outside the image gives $\mathcal{N}_i$. We use $r = 2$, giving at most 24 neighboring positions. For $\mathbf{v} \in \mathcal{N}_i$, let $z_{\mathbf{v}} = D(\mathbf{v})$ and let $\chi(\mathbf{v})$ indicate depth validity. It equals 1 when the depth satisfies the validity criterion of the experiment and 0 otherwise. Positions within the image that have invalid depth remain in $\mathcal{N}_i$ when computing observation coverage.

Fig. 2 illustrates the geometric construction. Fig. 2(a) shows the matched endpoints and their pixel neighborhoods in the source and target images. In each view, a reference plane parallel to the image plane is placed at the central depth z_i. After assigning this reference depth to the neighboring pixels, the lateral distance between $\mathbf{v}$ and the center is

$$r_{i\mathbf{v}} = z_i \sqrt{\left(\frac{v_x - u_i}{f_x}\right)^2 + \left(\frac{v_y - v_i}{f_y}\right)^2}.$$

Fig. 2(b) illustrates the relationship between lateral distance and pixel offset. Averaging the squared lateral distances over all neighboring positions within the image gives the squared lateral window scale

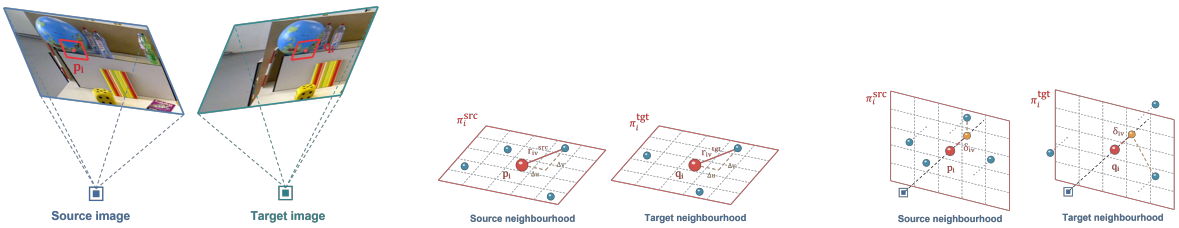
$$b_i^2 = \frac{z_i^2}{|\mathcal{N}_i|} \sum_{\mathbf{v} \in \mathcal{N}_i} \left[\left(\frac{v_x - u_i}{f_x}\right)^2 + \left(\frac{v_y - v_i}{f_y}\right)^2 \right], \tag{6}$$

where $\mathbf{u}_i = (u_i, v_i)^\top$, and f_x, f_y are focal lengths in pixels. Eq. (6) converts a fixed pixel window into a physical scale determined by depth and camera intrinsics, providing a reference for neighborhood depth differences.

Let the back-projection ray of the central pixel be

$$\begin{aligned} &\mathbf{a}_i = K^{-1}\widetilde{\mathbf{u}}_i, \\ &\alpha_i = \|\mathbf{a}_i\|_2^2 = 1 + \left(\frac{u_i - c_x}{f_x}\right)^2 + \left(\frac{v_i - c_y}{f_y}\right)^2. \end{aligned} \tag{7}$$

where (c_x, c_y) is the principal point and the intrinsics follow a zero-skew pinhole model.



(a) Corresponding points and local neighborhoods in the source and target views

(b) Lateral neighborhood distances in the reference plane

(c) Depth differences along the central ray

Fig. 2: Geometric construction of local depth support. (a) Matched endpoints and their local neighborhoods in the source and target views. (b) Lateral neighborhood distances in the reference plane. (c) Distances after mapping neighborhood depths onto the central ray. The orange point in (c) is an auxiliary position on the central ray with the same depth as the selected neighborhood point

In Fig. 2(c), mapping $z_{\mathbf{v}}$ onto the central ray gives the squared distance to the central point:

$$\delta_{i\mathbf{v}}^2 = \alpha_i (z_{\mathbf{v}} - z_i)^2. \tag{8}$$

The quantity $\delta_{i\mathbf{v}}$ describes the displacement obtained by mapping the depth difference along the central ray, while b_i describes the lateral window scale. Their ratio defines the support function

$$\kappa_{i\mathbf{v}} = \begin{cases} \dfrac{b_i^2}{b_i^2 + \delta_{i\mathbf{v}}^2}, & \chi(\mathbf{v}) = 1, \\ 0, & \chi(\mathbf{v}) = 0. \end{cases} \tag{9}$$

The contribution of a valid observation decreases monotonically with $\delta_{i\mathbf{v}}^2 / b_i^2$ and equals 1 when the depth difference is zero. Invalid observations contribute zero.

Averaging over all neighboring positions within the image gives the endpoint support

$$s_i = \frac{1}{|\mathcal{N}_i|} \sum_{\mathbf{v} \in \mathcal{N}_i} \kappa_{i\mathbf{v}}, \qquad 0 \le s_i \le 1. \tag{10}$$

Let $\mathcal{V}_i \subseteq \mathcal{N}_i$ be the set of positions with valid depth. If $|\mathcal{V}_i| > 0$, the expression can also be written as

$$s_i = \frac{|\mathcal{V}_i|}{|\mathcal{N}_i|} \left[\frac{1}{|\mathcal{V}_i|} \sum_{\mathbf{v} \in \mathcal{V}_i} \frac{b_i^2}{b_i^2 + \alpha_i (z_{\mathbf{v}} - z_i)^2} \right]. \tag{11}$$

Eq. (11) decomposes support into valid coverage and the mean consistency of valid observations. Missing depth reduces valid coverage, while inconsistent depth reduces the average contribution of valid observations. If $\mathcal{V}_i = \varnothing$, then $s_i = 0$.

This construction connects depth differences and lateral scale through a dimensionless ratio. If the set of valid positions is unchanged, a length-unit change $z \mapsto \lambda z$, with $\lambda > 0$, multiplies both b_i^2 and $\delta_{i\mathbf{v}}^2$ by λ^2, leaving s_i unchanged.

For correspondence c_i, compute endpoint supports s_i^s and s_i^t in the source and target views, respectively. Define correspondence-level depth support as

$$d_i = \min(s_i^s, s_i^t). \tag{12}$$

Because the correspondence constraint depends on the 3D positions of both endpoints, the endpoint with lower support determines d_i.

3.3 Depth-support-guided sampling and candidate prechecking

Assume that at least three correspondences have positive support. Normalizing d_i gives the probability distribution for the first draw:

$$\pi_i = \frac{d_i}{\sum_{j=1}^{N} d_j}. \tag{13}$$

To generate a rigid pose hypothesis, each round draws three distinct correspondences from $\mathcal{C}$. Let $\mathcal{S}_{k-1}$ contain the indices selected in the first $k-1$

draws of the current round, with $\mathcal{S}_0 = \varnothing$. For correspondences not yet selected, use the conditional distribution

$$\Pr(i_k = i \mid \mathcal{S}_{k-1}) = \frac{d_i}{\sum_{j \notin \mathcal{S}_{k-1}} d_j}, \quad i \notin \mathcal{S}_{k-1}, \quad k = 1, 2, 3. \tag{14}$$

Eq. (14) renormalizes the remaining weights after each selection so that the three indices are distinct. The next round samples again from the full candidate set. If fewer than three correspondences have positive support, a minimal triplet cannot be generated and estimation fails.

Sampling weights are computed from endpoint neighborhoods, but a minimal set must also satisfy the geometric constraints of a common rigid motion. For any a, b in $\mathcal{S} = \{i_1, i_2, i_3\}$, rigid distance preservation requires noisy correct correspondences to approximately satisfy

$$\|\mathbf{p}_a - \mathbf{p}_b\|_2 \approx \|\mathbf{q}_a - \mathbf{q}_b\|_2. \tag{15}$$

Define the edge-length inconsistency as

$$\Delta_{ab} = \left| \|\mathbf{p}_a - \mathbf{p}_b\|_2 - \|\mathbf{q}_a - \mathbf{q}_b\|_2 \right|. \tag{16}$$

We relate the precheck threshold to the 3D correspondence residual threshold τ. Suppose a common rigid transformation makes the residuals of correspondences a and b no greater than τ. Rigid distance preservation, Eq. (4), and the reverse triangle inequality give

$$\begin{aligned} \Delta_{ab} &= \left| \|R(\mathbf{p}_a - \mathbf{p}_b)\|_2 - \|\mathbf{q}_a - \mathbf{q}_b\|_2 \right| \\ &\le \|\boldsymbol{\epsilon}_a - \boldsymbol{\epsilon}_b\|_2 \\ &\le \|\boldsymbol{\epsilon}_a\|_2 + \|\boldsymbol{\epsilon}_b\|_2 \le 2\tau. \end{aligned} \tag{17}$$

We therefore require all three edges to satisfy

$$\max_{a,b \in \mathcal{S},\ a<b} \Delta_{ab} \le 2\tau. \tag{18}$$

If any edge has $\Delta_{ab} > 2\tau$, its two correspondences cannot both satisfy the residual threshold under a common rigid transformation, so the minimal set is rejected. Passing this check is a necessary condition for the three correspondences to satisfy the threshold jointly. Retained candidates proceed to degeneracy checking, pose estimation, and evaluation over all correspondences. The implementation adds a numerical tolerance of 10^{-9} mm when comparing edge lengths.

Rejected candidates skip three-point SVD and verification over all correspondences. A fixed attempt budget counts minimal-set draws, so a precheck rejection also consumes one attempt.

3.4 Rigid hypothesis estimation and evaluation

A minimal set that passes the geometric precheck generates a rigid pose hypothesis. For an index set $\mathcal{A}$, we use equally weighted least-squares rigid estimation [44]:

$$(R_{\mathcal{A}}, \mathbf{t}_{\mathcal{A}}) = \underset{R \in \mathrm{SO}(3),\, \mathbf{t} \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{i \in \mathcal{A}} \|R\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2^2. \tag{19}$$

For hypothesis generation, $\mathcal{A} = \mathcal{S}$, so each minimal set contains three correspondences. Let $m = |\mathcal{A}|$. The centroids of the two point sets are

$$\overline{\mathbf{p}} = \frac{1}{m} \sum_{i \in \mathcal{A}} \mathbf{p}_i, \qquad \overline{\mathbf{q}} = \frac{1}{m} \sum_{i \in \mathcal{A}} \mathbf{q}_i. \tag{20}$$

Center the point sets as $\widehat{\mathbf{p}}_i = \mathbf{p}_i - \overline{\mathbf{p}}$ and $\widehat{\mathbf{q}}_i = \mathbf{q}_i - \overline{\mathbf{q}}$, and construct the cross-covariance matrix

$$H_{\mathcal{A}} = \frac{1}{m} \sum_{i \in \mathcal{A}} \widehat{\mathbf{p}}_i \widehat{\mathbf{q}}_i^{\top}. \tag{21}$$

A triplet that passes the distance-consistency precheck may still contain collinear or coincident points. We therefore check spatial degeneracy before solving. Let $P_{\mathcal{A}}$ and $Q_{\mathcal{A}}$ have the centered source and target points as their rows, respectively. We require

$$\min\{\operatorname{rank}(P_{\mathcal{A}}), \operatorname{rank}(Q_{\mathcal{A}}), \operatorname{rank}(H_{\mathcal{A}})\} \ge 2. \tag{22}$$

The implementation determines numerical ranks from singular values, rejecting candidates with collinear or coincident points or insufficient cross-covariance rank. A failed check terminates the current hypothesis computation.

Singular value decomposition of $H_{\mathcal{A}}$ gives

$$
\begin{aligned}
H_{\mathcal{A}} &= U\Sigma V^{\top}, \\
R_{\mathcal{A}} &= V\,\mathrm{diag}\big(1, 1, \det(VU^{\top})\big)U^{\top}, \\
\mathbf{t}_{\mathcal{A}} &= \overline{\mathbf{q}} - R_{\mathcal{A}}\overline{\mathbf{p}}.
\end{aligned}
\tag{23}
$$

The determinant correction ensures that the resulting matrix belongs to SO(3), producing an orientation-preserving rigid transformation.

For each valid hypothesis $(R_h, \mathbf{t}_h)$, compute residuals $e_i^h = e_i(R_h, \mathbf{t}_h)$ over all correspondences and use truncated linear residual scoring [5]:

$$
S_h = \sum_{i=1}^{N} \phi_\tau(e_i^h), \qquad \phi_\tau(e) = \max\left(0, 1 - \frac{e}{\tau}\right). \tag{24}
$$

Eq. (24) accumulates continuous support from correspondences within the threshold; correspondences outside it contribute zero. It therefore uses both the number of supporting correspondences and their residual magnitudes. For a fixed correspondence set,

$$
S_h = N - \frac{1}{\tau}\sum_{i=1}^{N} \min(e_i^h, \tau). \tag{25}
$$

Maximizing S_h is equivalent to minimizing the sum of truncated absolute residuals. The residual cost of each correspondence is capped at τ, while residual differences within the threshold remain in the linear term.

The inlier set of a hypothesis is

$$
\mathcal{I}_h = \{i \mid e_i^h \le \tau\}. \tag{26}
$$

Among valid hypotheses with $|\mathcal{I}_h| \ge 3$, select the one with the highest score. Ties retain the earlier hypothesis. Replace $\mathcal{A}$ with its inlier set $\mathcal{I}_h$ and repeat the equally weighted estimation in Eqs. (19)–(23) to obtain the final pose. Estimation fails if no usable hypothesis is generated or if the final inlier set fails the degeneracy check. Final refinement is performed once on the fixed selected inlier set.

Algorithm 1 Depth-Support-Guided RANSAC (DSG-RANSAC)

Require: Pixel correspondences $\mathcal{M}$; depth maps D_s, D_t; intrinsics K_s, K_t; neighborhood radius r; residual threshold τ; attempt budget K

Ensure: Source-to-target rigid transformation $(R, \mathbf{t})$, or failure

Back-project correspondences with valid depth at both endpoints to construct $\mathcal{C}$
Compute depth support d_i using Eqs. (6)–(12)
if fewer than three d_i are positive **then return** failure
Initialize $\mathcal{H} \leftarrow \varnothing$
for $k = 1, \ldots, K$ **do**
    Draw a triplet $\mathcal{S}$ without replacement using Eq. (14)
    if any edge violates Eq. (18) **then continue**
    Construct centered point sets and their cross-covariance matrix
    if the degeneracy check in Eq. (22) fails **then continue**
    Generate $(R_h, \mathbf{t}_h)$ using Eq. (23)
    Evaluate residuals, inlier count, and the score in Eq. (24) over $\mathcal{C}$
    if $|\mathcal{I}_h| \ge 3$ **then** add the hypothesis and score to $\mathcal{H}$
end for
if $\mathcal{H} = \varnothing$ **then return** failure
Select the highest-scoring hypothesis; ties retain the earlier hypothesis
Perform one equally weighted SVD refinement on its inlier set
if refinement fails **then return** failure; **else return** $(R, \mathbf{t})$

3.5 Complete algorithm and computational cost

Algorithm 1 summarizes estimation under a fixed attempt budget. We analyze precomputation, candidate generation, and verification over all correspondences below.

Suppose each endpoint neighborhood contains at most M positions. Of K sampled minimal sets, let K_v pass the precheck and degeneracy checks and proceed to verification over all correspondences. Support computation costs $O(NM)$.

Each candidate checks only three edges and solves a fixed-dimensional rigid transformation, so geometric prechecking and three-point SVD each require constant work. Verification of one valid hypothesis costs $O(N)$, and final inlier refinement costs at most $O(N)$.

Let $C_{\mathrm{sam}}(N)$ denote the cost of one weighted triplet draw without replacement. The total time complexity is

$$O(NM + KC_{\mathrm{sam}}(N) + K + K_v N + N)\,. \quad (27)$$

The current implementation normalizes array weights and samples from them at a cost of $O(N)$ per draw. With a fixed neighborhood size, the overall worst-case complexity is therefore $O(KN)$. Prechecking reduces verification work by reducing K_v. Its benefit depends on the rejection rate, correspondence count, and actual costs of sampling and verification.

The implementation solves and verifies candidates in batches and retains candidate transformations and scores for final selection. Correspondence data require $O(N)$ space, expanded depth neighborhoods require at most $O(NM)$, and candidate records require $O(K)$. With a fixed neighborhood size, total space complexity is $O(N + K)$.

To examine the effect of computational budgets, experiments use both fixed-attempt and fixed-time stopping rules. The former counts attempts as in Algorithm 1. The latter uses elapsed backend time and recent batch costs to decide whether to generate more candidates, reserving time for final selection and refinement. Depth support computation is included in the time budget. Batch boundaries and final refinement may introduce small timing deviations. Actual total time and its deviation from the target budget are recorded; batch settings are described in the experimental setup.

4 Experiments and analysis

4.1 Experimental setup

Datasets and correspondence construction. We use TUM RGB-D indoor data [45] and ROBI industrial part data [46], with the test scope listed in Table 1. TUM includes adjacent frame pairs from six sequences. ROBI includes adjacent view pairs from Scenes 4, 5, and 8 for seven part categories, giving 21 scenes. These datasets assess registration under indoor and industrial part observation conditions, respectively.

The experiments with different frontends use the complete lists of adjacent view pairs.

Table 1: Datasets and test scope

Dataset	Test scope	Frame pairs
TUM	Adjacent frame pairs from six sequences	8357
ROBI	Seven part categories (Scenes 4, 5, and 8)	2386

Correspondence construction. Efficient LoFTR is the main frontend. Its pixel matches are filtered for valid depth and back-projected into 3D correspondences. SuperPoint with brute-force matching (BF) and SIFT provide additional correspondence sources. In comparisons using the same frontend, conventional RANSAC and our method use the same cached correspondences. Depth lookup and back-projection follow the experimental settings of each frontend.

Methods and parameters. Conventional RANSAC uses uniform triplet sampling and inlier-count scoring. Our method uses depth-support-guided sampling, three-edge geometric prechecking, and truncated linear scoring. Comparisons with existing estimators include PROSAC(q), GCC, and SC^2-PCR, where q denotes frontend matching quality. An additional experiment uses the same PROSAC schedule with rankings based on depth support d and frontend quality q to assess their effects on sampling and pose estimation. PROSAC uses an existing schedule adaptation, truncated linear scoring, and equally weighted inlier refinement. GCC uses its native estimation pipeline, and SC^2-PCR uses the authors' core solver.

Fixed-attempt random sampling experiments use seeds 42–46. Sampling diagnostics for the main frontend draw 500 triplet candidates per frame pair and seed. The depth neighborhood radius is 2, corresponding to a 5×5 pixel window around the center.

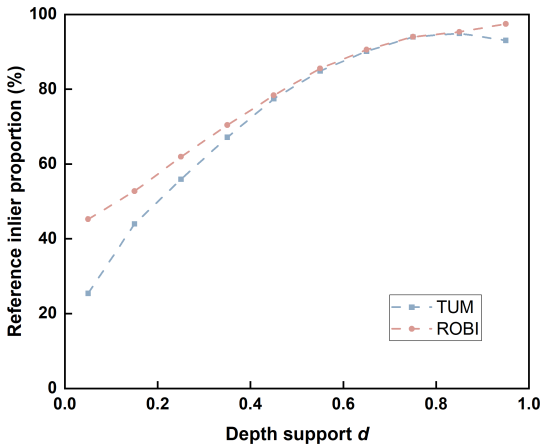
Evaluation metrics. Registration error is computed from the relative error between the estimated and reference transformations. Let T be the estimated source-to-target transformation and G the reference transformation. The adjacent-view experiments with the main frontend use $E = GT^{-1}$. Rotation error is the rotation angle of the error transformation in degrees, and translation error is the norm of its translation vector in mm. Root mean square errors (RMSEs) are computed across frame pairs and then aggregated across random seeds according to each experiment's protocol. Error conventions and statistical aggregation for other experiments are specified with the results.

Sampling quality is evaluated by the all-inlier triplet rate, the all-inlier and nondegenerate triplet rate, and the qualified hypothesis rate. The first two measure the proportions of sampled triplets whose three correspondences are all reference geometric inliers and whose triplets additionally pass the degeneracy check, respectively. The qualified hypothesis rate is the proportion of all candidates that directly generate valid three-point hypotheses meeting the accuracy threshold. Hypothesis error is the RMS difference between the actions of the estimated and reference transformations on at most 1000 fixed depth evaluation points per source view. Both the 3D residual threshold for reference geometric inliers and the hypothesis qualification threshold are 30 mm on TUM and 3 mm on ROBI.

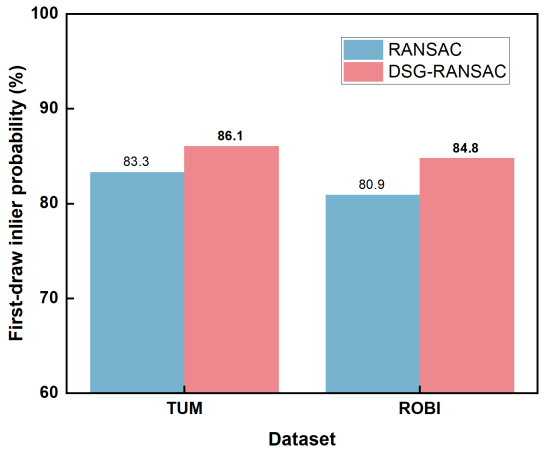
4.2 Comparisons with the same frontend

4.2.1 Depth support and sampling quality

We first examine the relationship between depth support d and reference geometric inliers, and the effect of d-guided sampling on the inlier probability of the first draw. Fig. 3 uses the complete Efficient LoFTR correspondence caches for adjacent views in TUM and ROBI. It shows reference inlier ratios across support intervals and first-draw inlier probabilities under uniform and d-guided sampling.



(a) Depth support and reference inlier ratio



(b) Reference inlier probability of the first correspondence draw

Fig. 3: Statistical relationship between depth support and reference geometric inliers. (a) Correspondences are grouped into support intervals of width 0.1, and the reference inlier ratio is computed over all correspondences in each interval. (b) The probability that the first drawn correspondence is a reference inlier is computed per frame pair and then averaged with equal weight across pairs

Fig. 3(a) shows that the reference inlier ratio generally increases with depth support on both datasets. For pairs containing both reference inliers and outliers, we compute the mean support of each group per pair and then average equally across pairs. The mean supports of reference inliers and outliers are 0.6009 and 0.4602 on TUM, and 0.6412 and 0.4425 on ROBI, respectively; the latter includes 2385 pairs.

In Fig. 3(b), the first-draw inlier probability increases from 83.28% under uniform sampling to 86.05% under d-guided sampling on TUM, and from 80.92% to 84.80% on ROBI. The reported gains are 2.78 and 3.88 percentage points, respectively. Depth support is statistically associated with reference inliers.

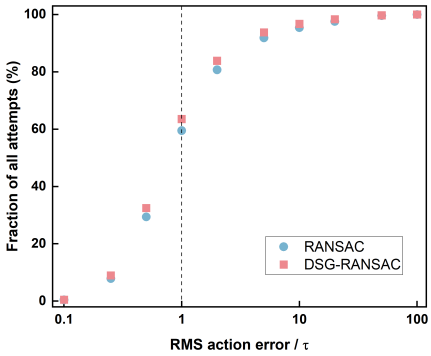
Table 2 examines the effect of depth support on triplet sampling by comparing candidate quality under uniform and d-guided sampling. All three metrics are computed before geometric prechecking.

Table 2: Candidate quality under different sampling strategies (%)

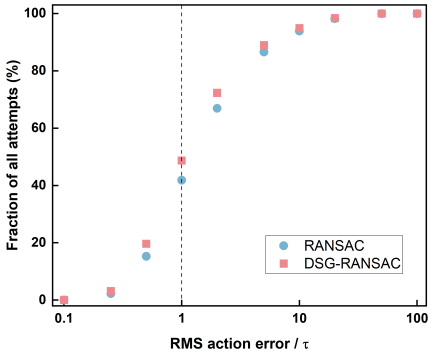
Metric	TUM		ROBI	
	Uniform	d-guided	Uniform	d-guided
All-inlier triplet rate	60.34	**66.25**	57.78	**66.33**
All-inlier and nondegenerate triplet rate	60.34	**66.25**	57.78	**66.32**
Qualified hypothesis rate	59.52	**63.52**	41.87	**48.72**

Table 2 shows that d-guided sampling increases both the all-inlier triplet rate and the qualified hypothesis rate. On TUM, these rates increase from 60.34% to 66.25% and from 59.52% to 63.52%, respectively. On ROBI, they increase from 57.78% to 66.33% and from 41.87% to 48.72%. For both strategies, the all-inlier and nondegenerate triplet rate is close to the all-inlier triplet rate.

Fig. 4 further shows the proportions of raw three-point hypotheses within different error ranges, extending the candidate-quality analysis beyond the qualification threshold in Table 2.



(a) Raw hypothesis error on TUM



(b) Raw hypothesis error on ROBI

Fig. 4: Cumulative proportions of raw three-point hypothesis errors. (a) TUM. (b) ROBI. The horizontal axis is the ratio of RMS action error to threshold τ. The vertical axis is the proportion of all sampling attempts that produce valid hypotheses within the corresponding error range

Within an error of 0.5τ, Fig. 4 gives cumulative proportions of 29.38% and 32.44% for uniform and d-guided sampling on TUM, and 15.28% and 19.67% on ROBI. At 2τ, the corresponding proportions are 80.72% and 83.86% on TUM, and 66.91% and 72.32% on ROBI. Guided sampling produces a higher cumulative proportion of valid hypotheses at both error levels. Together, Fig. 3, Table 2, and Fig. 4 show a statistical association between depth support and reference geometric inliers, and increased proportions of all-inlier triplets and valid pose hypotheses under support-guided sampling.

4.2.2 Registration errors on TUM and ROBI

Table 3 evaluates the effect of depth-support-guided sampling on final pose estimation with the same image matching frontend. Both estimators use identical 3D correspondences. Rotation and translation errors are computed for six TUM sequences and seven ROBI part categories.

Table 3: Registration errors with the same frontend (rotation RMSE / translation RMSE, ° / mm)

(a) TUM

Sequence	Frame pairs	RANSAC	DSG-RANSAC
fr1/desk	595	0.515 / 10.70	**0.502** / **10.30**
fr1/room	1361	0.379 / 7.07	**0.372** / **6.62**
fr1/xyz	794	0.325 / 5.49	**0.317** / **5.26**
fr2/desk	2206	0.290 / 5.16	**0.284** / **4.70**
fr3/struct	1011	0.545 / 11.46	**0.514** / **10.65**
fr3/office	2390	0.281 / 6.50	**0.267** / **5.75**

(b) ROBI

Part	Frame pairs	RANSAC	DSG-RANSAC
Zigzag	341	1.128 / 7.57	**0.972** / **6.54**
Chrome screw	341	1.450 / 8.02	**0.349** / **2.44**
Gear	340	2.075 / 13.70	**1.381** / **8.68**
Eye bolt	341	1.125 / 6.22	**1.074** / **5.96**
Tube fitting	341	1.198 / 7.36	**0.722** / **4.45**
DIN	341	1.403 / 9.27	**0.975** / **6.51**
DSub	341	0.931 / 5.63	**0.694** / **4.13**

DSG-RANSAC has lower rotation and translation error values than RANSAC on all six TUM sequences. On fr1/desk, the errors are 0.515° / 10.70 mm for RANSAC and 0.502° / 10.30 mm for DSG-RANSAC. On fr3/office, they are 0.281° / 6.50 mm and 0.267° / 5.75 mm, respectively.

DSG-RANSAC also has lower rotation and translation error values on all seven ROBI part categories. For Chrome screw, rotation error decreases from 1.450° to 0.349° and translation error from 8.02 mm to 2.44 mm. For Gear, the errors decrease from 2.075° to 1.381° and from 13.70 mm to 8.68 mm.

With identical frontend correspondences, Table 3 shows lower rotation and translation errors for DSG-RANSAC across the six TUM sequences and seven ROBI categories. Together with the sampling-quality results, these findings show that depth-support-guided sampling generates more high-quality candidates and yields more accurate registration in indoor and industrial part scenes.

4.2.3 Registration errors under equal time budgets

Table 4 compares registration errors under equal time budgets. Conventional RANSAC draws 500 samples per frame pair. Its measured backend time for the same pair and seed becomes the DSG-RANSAC budget. The mean budgets are approximately 75 ms on TUM and 60 ms on ROBI. DSG-RANSAC time includes depth support computation and final pose refinement.

Table 4: Registration errors under equal time budgets (rotation RMSE / translation RMSE, ° / mm). Mean attempts refer to DSG-RANSAC.

(a) TUM

Sequence	RANSAC	DSG-RANSAC	Mean attempts
fr1/desk	0.528 / 11.00	**0.521** / **10.84**	153
fr1/room	0.380 / 7.10	**0.377** / **6.88**	166
fr1/xyz	0.329 / 5.60	**0.321** / **5.40**	171
fr2/desk	0.290 / 5.07	**0.288** / **4.72**	150
fr3/struct	0.544 / 11.53	**0.517** / **10.79**	161
fr3/office	0.274 / 6.41	**0.268** / **5.89**	161

(b) ROBI

Part	RANSAC	DSG-RANSAC	Mean attempts
Zigzag	6.943 / 39.24	**6.358** / **34.25**	305
Chrome screw	1.343 / 7.92	**1.185** / **3.00**	251
Gear	9.326 / 46.21	**8.117** / **43.77**	292
Eye bolt	**1.102** / 6.08	1.112 / **5.46**	262
Tube fitting	**5.941** / **37.43**	5.985 / 37.89	279
DIN	1.337 / 8.83	**1.221** / **7.99**	262
DSub	0.897 / 5.41	**0.839** / **4.99**	241

Across jointly valid runs on all six TUM sequences, DSG-RANSAC lowers both errors relative to RANSAC with 150–171 mean attempts.

Table 5: Registration errors with different frontends (rotation RMSE / translation RMSE, ° / mm)

(a) TUM

Sequence	Efficient LoFTR		SuperPoint + BF		SIFT	
	RANSAC	DSG-RANSAC	RANSAC	DSG-RANSAC	RANSAC	DSG-RANSAC
fr1/desk	0.515 / 10.70	**0.502** / **10.30**	0.535 / 11.26	**0.517** / **10.77**	0.555 / 11.92	**0.539** / **11.24**
fr1/room	0.379 / 7.07	**0.372** / **6.62**	0.379 / 7.29	**0.372** / **6.69**	0.394 / **8.54**	**0.391** / 8.72
fr1/xyz	0.325 / 5.49	**0.317** / **5.26**	0.339 / 5.74	**0.332** / **5.46**	0.365 / 6.41	**0.355** / **6.08**
fr2/desk	0.290 / 5.16	**0.284** / **4.70**	0.276 / 4.11	**0.273** / **3.83**	0.282 / 4.45	**0.277** / **3.91**
fr3/struct	0.545 / 11.46	**0.514** / **10.65**	0.593 / 13.67	**0.569** / **12.98**	0.653 / 15.02	**0.592** / **13.74**
fr3/office	0.281 / 6.50	**0.267** / **5.75**	0.270 / 6.42	**0.257** / **5.68**	0.293 / 7.89	**0.275** / **6.86**

(b) ROBI

Part category	Efficient LoFTR		SuperPoint + BF		SIFT	
	RANSAC	DSG-RANSAC	RANSAC	DSG-RANSAC	RANSAC	DSG-RANSAC
Zigzag	1.128 / 7.57	**0.972** / **6.54**	**13.120** / **61.04**	13.421 / 64.12	9.672 / **43.31**	**9.281** / 43.77
Chrome screw	1.450 / 8.02	**0.349** / **2.44**	9.398 / 46.92	**5.629** / **35.88**	16.823 / 78.14	**13.699** / **66.25**
Gear	2.075 / 13.70	**1.381** / **8.68**	12.488 / 58.78	**11.146** / **31.45**	22.694 / 97.81	**21.815** / **90.16**
Eye bolt	1.125 / 6.22	**1.074** / **5.96**	5.644 / 34.07	**2.287** / **15.02**	15.270 / **57.85**	**15.264** / 57.93
Tube fitting	1.198 / 7.36	**0.722** / **4.45**	3.738 / 21.17	**3.192** / **18.66**	20.825 / 91.37	**19.747** / **89.12**
DIN	1.403 / 9.27	**0.975** / **6.51**	5.637 / 30.68	**5.511** / **29.26**	24.194 / 99.62	**20.313** / **80.64**
DSub	0.931 / 5.63	**0.694** / **4.13**	**6.656** / **34.17**	7.187 / 39.56	20.624 / 73.46	**18.129** / **63.40**

On ROBI, DSG-RANSAC has lower rotation and translation errors for five categories: Zigzag, Chrome screw, Gear, DIN connector, and DSub connector. Errors decrease from 1.343° / 7.92 mm to 1.185° / 3.00 mm for Chrome screw, and from 1.337° / 8.83 mm to 1.221° / 7.99 mm for DIN connector. For Eye bolt, translation error decreases from 6.08 mm to 5.46 mm, while rotation error increases from 1.102° to 1.112°. Mean DSG-RANSAC attempts range from 241 to 305 in this ROBI experiment.

Under equal time budgets, DSG-RANSAC achieves lower rotation and translation errors on all six TUM sequences and five ROBI categories. Its 150–171 mean attempts on TUM indicate that depth-support guidance can obtain better registration with fewer candidate attempts. Performance on ROBI varies with part category: Eye bolt has lower translation error but slightly higher rotation error, and Tube fitting has slightly higher errors. These results show the potential of depth-support guidance to improve registration accuracy in indoor scenes and some industrial part scenes.

4.3 Experiments with different frontends

To examine applicability across correspondence sources, Table 5 compares conventional RANSAC and DSG-RANSAC using Efficient LoFTR, SuperPoint + BF, and SIFT.

Across the six TUM sequences, DSG-RANSAC has lower rotation error values than conventional RANSAC with all three frontends. Translation error values are lower on all sequences with SuperPoint + BF and Efficient LoFTR, and on five sequences with SIFT. On fr3/struct, for example, SuperPoint + BF errors decrease from 0.593° and 13.67 mm to 0.569° and 12.98 mm; SIFT errors decrease from 0.653° and 15.02 mm to 0.592° and 13.74 mm. For SIFT on fr1/room, the displayed translation error increases from 8.54 mm to 8.72 mm.

ROBI results vary with frontend and part category. With SuperPoint + BF, Chrome screw errors decrease from 9.398° / 46.92 mm to 5.629° / 35.88 mm, while DSub connector errors increase from 6.656° / 34.17 mm to 7.187° / 39.56 mm. With SIFT, DIN connector errors decrease from

24.194° / 99.62 mm to 20.313° / 80.64 mm, while Zigzag translation error increases from 43.31 mm to 43.77 mm. Efficient LoFTR yields lower values for both errors across all seven categories.

Table 5 shows that DSG-RANSAC can use correspondences from learned and conventional feature frontends and achieve lower registration error values under different input conditions. On TUM, rotation errors are lower across all six sequences with all three frontends, and translation errors are lower on most sequences. On ROBI, lower errors occur for several industrial parts, including Chrome screw with SuperPoint + BF and DIN connector with SIFT. Depth-support guidance can therefore be used for correspondence sampling with different frontends and has the potential to improve registration accuracy in indoor and multiple industrial part scenes.

4.4 Comparison with existing estimators

Table 6: Registration errors of different estimators (rotation RMSE / translation RMSE, ° / mm)

(a) TUM

Sequence	DSG-RANSAC	PROSAC (q)	GCC	SC^2-PCR
fr1/desk	0.502/10.30	0.506/10.34	0.587/10.98	**0.497/10.14**
fr1/room	0.372/6.62	0.372/6.67	0.449/9.91	**0.370/6.44**
fr1/xyz	0.317/5.26	0.319/5.28	0.411/6.99	**0.315/5.17**
fr2/desk	**0.284**/4.70	**0.284**/4.69	0.335/7.28	**0.284/4.61**
fr3/struc	**0.514/10.65**	0.521/10.81	0.594/12.40	0.519/10.67
fr3/office	**0.267/5.75**	0.268/5.76	0.363/9.12	0.268/5.76

(b) ROBI

Part	DSG-RANSAC	PROSAC (q)	GCC	SC^2-PCR
Zigzag	**0.972/6.54**	7.341/38.97	9.422/47.79	5.837/34.01
Chrome screw	**0.349/2.44**	2.094/12.18	4.446/25.13	1.211/7.19
Gear	**1.381/8.68**	7.923/43.45	7.748/40.37	2.045/13.19
Eye bolt	**1.074/5.96**	1.171/6.20	3.184/19.12	1.081/5.97
Tube fitting	**0.722/4.45**	6.898/42.94	8.408/47.54	5.277/32.93
DIN	**0.975/6.51**	1.243/8.04	2.895/17.56	1.185/7.72
DSub	**0.694/4.13**	0.850/5.06	2.561/15.25	0.801/4.77

Table 6 compares DSG-RANSAC with PROSAC(q), GCC, and SC^2-PCR. All methods use Efficient LoFTR correspondences. PROSAC(q) ranks correspondences by frontend quality, while GCC and SC^2-PCR use their respective geometric estimation pipelines.

On TUM, DSG-RANSAC and PROSAC(q) have similar error values, with both errors lower than GCC across all six sequences. On fr3/struct, DSG-RANSAC, PROSAC(q), and GCC achieve 0.514° / 10.65 mm, 0.521° / 10.81 mm, and 0.594° / 12.40 mm, respectively. SC^2-PCR has lower error values on some sequences, including 0.497° / 10.14 mm on fr1/desk.

On ROBI, the DSG-RANSAC column has lower error values than the three comparison methods. For Chrome screw, the errors are 0.349° / 2.44 mm for DSG-RANSAC, 2.094° / 12.18 mm for PROSAC(q), 4.446° / 25.13 mm for GCC, and 1.211° / 7.19 mm for SC^2-PCR.

Table 6 shows that DSG-RANSAC has errors close to PROSAC(q) and lower than GCC on the six TUM sequences. Its displayed rotation and translation errors are lower on the seven ROBI categories. These results characterize depth-support-guided registration in indoor and industrial part scenes.

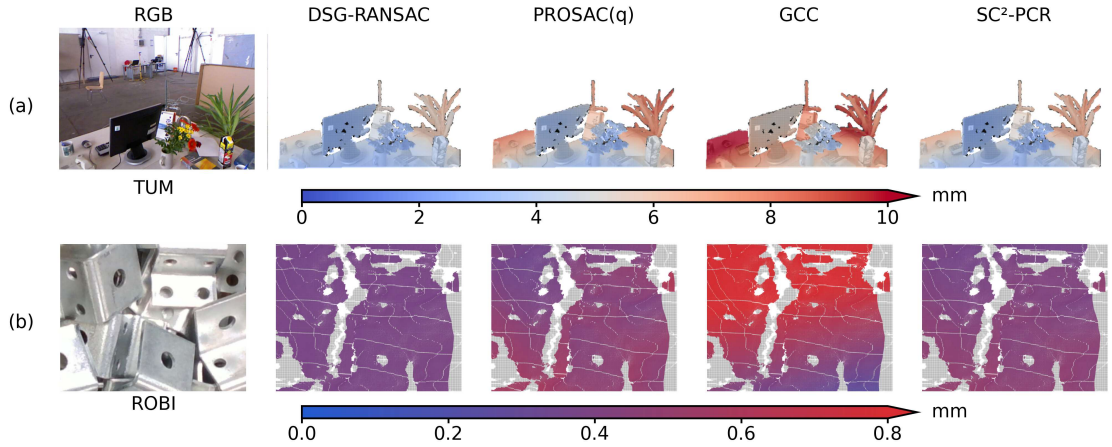


Fig. 5: Spatial errors of registration results on TUM and ROBI. (a) TUM. (b) ROBI. Gray denotes the source point cloud; color indicates the positional difference between target points transformed by the estimated and reference transformations

Fig. 5 presents registration results for two frame pairs from TUM and ROBI, illustrating the spatial errors of different estimators in specific scenes.

In the TUM pair in Fig. 5(a), low-displacement regions in the DSG-RANSAC result cover the point cloud around the monitor and desk. GCC shows more pronounced high-displacement regions on the left side of the desk and around the plant.

In the ROBI pair in Fig. 5(b), the displayed DSG-RANSAC result is dominated by low-displacement regions, whereas GCC exhibits a more extensive high-displacement region.

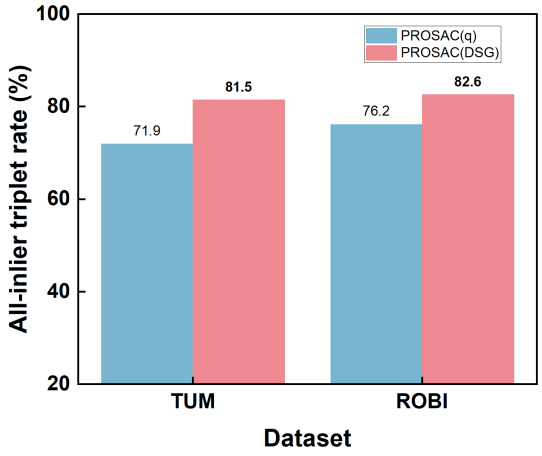
4.5 Depth support versus frontend quality as a prior

To examine local depth support in another sampling framework, we use it to rank correspondences for PROSAC and compare it with frontend matching quality ranking. Both variants use identical Efficient LoFTR correspondences, the same PROSAC schedule, and a budget of 500 sampling attempts. They share the same three-point SVD solver, truncated linear scoring, and final equally weighted inlier refinement. PROSAC(DSG) ranks by depth support d, while PROSAC(q) ranks by frontend matching quality q.

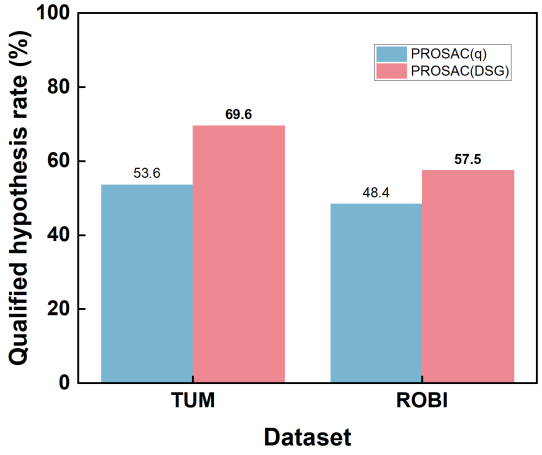
Fig. 6 compares all-inlier triplet rates and qualified hypothesis rates under the two ranking priors. Both metrics use all 500 sampling attempts per frame pair and seed as the denominator. Bar heights are means of the statistical proportions across seeds 42–46. TUM and ROBI contain 8357 and 2386 adjacent view pairs, respectively.

In Fig. 6(a), the TUM all-inlier triplet rate increases from 71.9% for PROSAC(q) to 81.5% for PROSAC(DSG), a gain of 9.6 percentage points. On ROBI, it increases from 76.2% to 82.6%, a gain of 6.4 percentage points. Depth support ranking enables PROSAC to draw more all-inlier triplets on both datasets.

In Fig. 6(b), the qualified hypothesis rate increases from 53.6% to 69.6% on TUM, a gain of 16.0 percentage points, and from 48.4% to 57.5% on ROBI, a gain of 9.1 percentage points. Under the same sampling schedule, depth support ranking increases the proportion of raw pose hypotheses meeting the reference accuracy threshold.



(a) All-inlier triplet rate



(b) Qualified hypothesis rate

Fig. 6: PROSAC sampling quality under different ranking priors. (a) All-inlier triplet rate. (b) Qualified hypothesis rate

These results show that local depth support transfers across sampling frameworks. The prior supports both our weighted sampling without replacement and PROSAC's progressive ranked sampling, increasing all-inlier triplet rates and qualified hypothesis rates on TUM and ROBI. Local depth support can be incorporated into different pose estimation frameworks as independent correspondence quality information, providing a transferable geometric prior for methods based on correspondence ranking or guided sampling.

5 Discussion

5.1 Local depth support and transfer across sampling frameworks

Local depth support organizes valid-depth coverage and depth variation around endpoints into correspondence-level information, allowing depth observations to inform minimal-set selection. Fig. 3 shows that higher-support correspondences have a higher reference geometric inlier ratio in aggregate. Table 2 and Fig. 4 further show that using this information to guide triplet sampling increases the proportions of all-inlier triplets and high-quality pose hypotheses. Individual correspondence support describes local observations. Whether three correspondences jointly produce an accurate pose also depends on their geometric distribution and spatial positions. First-draw inlier probability, all-inlier triplet rate, and qualified hypothesis rate therefore describe different stages of candidate generation.

Depth support can also serve as an independent ranking prior. Under the same PROSAC schedule in Fig. 6, depth support ranking produces higher all-inlier triplet rates and qualified hypothesis rates than frontend quality ranking on both TUM and ROBI. Together with the weighted sampling results in Sect. 4.2, these findings demonstrate the use of local depth support in two sampling frameworks. The prior also provides transferable geometric information for other pose estimators that rank correspondences or guide sampling by correspondence quality.

5.2 Performance across frontends and scenes

Depth support is computed from endpoint depth neighborhoods and can be used with different image matching frontends. In Table 5, DSG-RANSAC has lower rotation errors than conventional RANSAC on all six TUM sequences with Efficient LoFTR, SuperPoint + BF, and SIFT. ROBI results vary with frontend and part category. With SuperPoint + BF, both errors decrease for Chrome screw but increase for DSub connector. With SIFT, both decrease for DIN connector, while Zigzag translation error increases slightly.

Depth-support guidance can accommodate correspondences from different sources, while final registration gains depend on the input correspondences and scene conditions.

Table 6 shows similar errors for DSG-RANSAC and PROSAC(q) on TUM, with lower errors for SC^2-PCR on some sequences. DSG-RANSAC achieves lower errors in the displayed ROBI results. The methods use correspondence quality and geometric relationships differently, and their results reflect their estimation pipelines in the two scene types. Depth support supplies endpoint-neighborhood information that can be combined with inter-correspondence geometric constraints for candidate search and evaluation.

5.3 Accuracy and computational budgets

Depth support computation and guided sampling add work to candidate generation. Three-edge geometric prechecking allows candidates that violate edge-length consistency to skip subsequent pose estimation and verification over all correspondences. The overall timing effect depends on the actual costs of support computation, sampling, and verification, as well as the number of candidates rejected by prechecking.

Under the equal-time setting in Table 4, DSG-RANSAC makes 150–171 mean attempts on TUM, compared with 500 for conventional RANSAC, while achieving lower rotation and translation errors across jointly valid runs on all six sequences. On ROBI, both errors are lower for five categories; Eye bolt has slightly higher rotation error, and Tube fitting has slightly higher values for both errors. Under a limited time budget, higher candidate quality can therefore accompany fewer sampling attempts, while registration performance varies with scene and correspondence input.

6 Conclusion

We have proposed DSG-RANSAC for RGB-D coarse registration with unreliable correspondences. The method constructs correspondence-level local depth support from valid-depth coverage and normalized depth differences around matched endpoints. It uses this support to guide triplet sampling without replacement and combines three-edge geometric prechecking, rigid pose estimation, truncated linear scoring, and inlier-based refinement to obtain the registration transformation.

Experiments on TUM indoor scenes and ROBI industrial parts show a statistical association between local depth support and reference geometric inliers. Guided sampling increases the proportions of all-inlier triplets and qualified hypotheses. With the same frontend, DSG-RANSAC has lower rotation and translation errors than conventional RANSAC on all six TUM sequences and seven ROBI categories. Experiments with different frontends show that the method accommodates multiple correspondence sources, with performance varying across frontends and scenes. Under equal time budgets, it achieves lower rotation and translation errors on all six TUM sequences and five ROBI categories.

Using depth support to rank correspondences for PROSAC produces higher all-inlier triplet rates and qualified hypothesis rates than frontend quality ranking on both datasets. Together with weighted sampling without replacement, these experiments show improved candidate quality in two sampling frameworks and provide a transferable geometric prior for other pose estimation methods using correspondence quality.

Statements and Declarations

Funding

This work was supported by the National Natural Science Foundation of China (No. 12402217), the Fundamental Research Funds for the Central Universities (FRF-TP-26-046), and the National Key Laboratory of Metal Forming Technology and Heavy Equipment (No. SKLMF-2025-010).

Data Availability

The TUM RGB-D Dataset and Benchmark and the ROBI dataset analyzed in this study are publicly available from their respective project websites. The registration results are reported in the article.

Competing Interests

The authors declare that they have no competing interests relevant to the content of this article.

References

[1] Oğuz, E., Doğan, Y., Güdükbay, U., Karaşan, O., Pınar, M.: Point cloud registration with quantile assignment. Machine Vision and Applications 35(3), 38 (2024). https://doi.org/10.1007/s00138-024-01517-3

[2] Guo, Z., Ma, H.: Enhanced point cloud registration for workpieces using triangular constraint sampling consistency in complex industrial environment. The Visual Computer 42(5), 221 (2026). https://doi.org/10.1007/s00371-026-04421-1

[3] Wu, Z., Wang, Y., Mo, Y., Zhu, Q., Xie, H., Wu, H., Feng, M., Mian, A.: Multiview Point Cloud Registration Based on Minimum Potential Energy for Free-Form Blade Measurement. IEEE Transactions on Instrumentation and Measurement 71, 1–14 (2022). https://doi.org/10.1109/tim.2022.3169559

[4] 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). https://doi.org/10.1145/358669.358692

[5] Cheng, Y., Huang, Z., Quan, S., Cao, X., Zhang, S., Yang, J.: Sampling locally, hypothesis globally: accurate 3D point cloud registration with a RANSAC variant. Visual Intelligence 1(1), 20 (2023). https://doi.org/10.1007/s44267-023-00022-x

[6] Chum, O., Matas, J.: Matching with PROSAC — Progressive Sample Consensus. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 220–226 (2005). https://doi.org/10.1109/cvpr.2005.221

[7] 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 the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13221–13231 (2022). https://openaccess.thecvf.com/content/CVPR2022/html/Chen_SC2-PCR_A_Second_Order_Spatial_Compatibility_for_Efficient_and_Robust_CVPR_2022_paper.html

[8] Kumar, S., Chien, C.-H., Kimia, B.: Geometric Correspondence Consistency in RGB-D Relative Pose Estimation. In: International Conference on 3D Vision (3DV), pp. 469–479 (2025). https://doi.org/10.1109/3dv66043.2025.00048

[9] Khoshelham, K., Oude Elberink, S.: Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. Sensors 12(2), 1437–1454 (2012). https://doi.org/10.3390/s120201437

[10] Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992). https://doi.org/10.1109/34.121791

[11] Segal, A., Haehnel, D., Thrun, S.: Generalized-ICP. In: Robotics: Science and Systems V (2009). https://doi.org/10.15607/rss.2009.v.021

[12] Holz, D., Ichim, A.E., Tombari, F., Rusu, R.B., Behnke, S.: Registration with the Point Cloud Library: A Modular Framework for Aligning in 3-D. IEEE Robotics & Automation Magazine 22(4), 110–124 (2015). https://doi.org/10.1109/mra.2015.2432331

[13] Anderson, J.D., Raettig, R.M., Larson, J., Nykl, S.L., Taylor, C.N., Wischgoll, T.: Delaunay walk for fast nearest neighbor: accelerating correspondence matching for ICP. Machine Vision and Applications 33(2), 31 (2022). https://doi.org/10.1007/s00138-022-01279-w

[14] Bojanić, D., Bartol, K., Forest, J., Petković, T., Pribanić, T.: Addressing the generalization of 3D registration methods with a featureless baseline and an unbiased benchmark.

Machine Vision and Applications 35(3), 41 (2024). https://doi.org/10.1007/s00138-024-01510-w

[15] Slimani, K., Tamadazte, B., Achard, C.: Rocnet: 3D robust registration of points clouds using deep learning. Machine Vision and Applications 35(4), 100 (2024). https://doi.org/10.1007/s00138-024-01584-6

[16] Vodianyk, B., Nava-Baro, E., Popov, A.: SIGMAR-3D: scale-invariant guaranteed monotone accurate registration for 3D scenes. Machine Vision and Applications 38(1), 1 (2027). https://doi.org/10.1007/s00138-026-01941-7

[17] Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004). https://doi.org/10.1023/b:visi.0000029664.99615.94

[18] DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperPoint: Self-Supervised Interest Point Detection and Description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 224–236 (2018). https://doi.org/10.1109/cvprw.2018.00060

[19] Wang, Y., He, X., Peng, S., Tan, D., Zhou, X.: Efficient LoFTR: Semi-Dense Local Feature Matching with Sparse-Like Speed. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 21666–21675 (2024). https://openaccess.thecvf.com/content/CVPR2024/html/Wang_Efficient_LoFTR_Semi-Dense_Local_Feature_Matching_with_Sparse-Like_Speed_CVPR_2024_paper.html

[20] Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperGlue: Learning Feature Matching With Graph Neural Networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4938–4947 (2020). https://doi.org/10.1109/cvpr42600.2020.00499

[21] Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: LoFTR: Detector-Free Local Feature Matching With Transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8922–8931 (2021). https://doi.org/10.1109/cvpr46437.2021.00881

[22] Lindenberger, P., Sarlin, P.-E., Pollefeys, M.: LightGlue: Local Feature Matching at Light Speed. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 17627–17638 (2023). https://openaccess.thecvf.com/content/ICCV2023/html/Lindenberger_LightGlue_Local_Feature_Matching_at_Light_Speed_ICCV_2023_paper.html

[23] Wang, S., Kannala, J., Pollefeys, M., Barath, D.: Guiding Local Feature Matching with Surface Curvature. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 17981–17991 (2023). https://doi.org/10.1109/iccv51070.2023.01648

[24] Zeng, A., Song, S., Niessner, M., Fisher, M., Xiao, J., Funkhouser, T.: 3DMatch: Learning Local Geometric Descriptors From RGB-D Reconstructions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1802–1811 (2017). https://doi.org/10.1109/cvpr.2017.29

[25] Choy, C., Park, J., Koltun, V.: Fully Convolutional Geometric Features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8958–8966 (2019). https://openaccess.thecvf.com/content_ICCV_2019/html/Choy_Fully_Convolutional_Geometric_Features_ICCV_2019_paper.html

[26] Bai, X., Luo, Z., Zhou, L., Fu, H., Quan, L., Tai, C.-L.: D3Feat: Joint Learning of Dense Detection and Description of 3D Local Features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6359–6367 (2020). https://doi.org/10.1109/cvpr42600.2020.00639

[27] Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., Schindler, K.: Predator: Registration of 3D Point Clouds With Low Overlap. In: Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition, pp. 4267–4276 (2021). https://openaccess.thecvf.com/content/CVPR2021/html/Huang_Predator_Registration_of_3D_Point_Clouds_With_Low_Overlap_CVPR_2021_paper.html

[28] Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Xu, K.: Geometric Transformer for Fast and Robust Point Cloud Registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11143–11152 (2022). https://doi.org/10.1109/cvpr52688.2022.01086

[29] Yew, Z.J., Lee, G.H.: REGTR: End-to-End Point Cloud Correspondences With Transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6677–6686 (2022). https://doi.org/10.1109/cvpr52688.2022.00656

[30] Park, J., Zhou, Q.-Y., Koltun, V.: Colored Point Cloud Registration Revisited. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 143–152 (2017). https://openaccess.thecvf.com/content_iccv_2017/html/Park_Colored_Point_Cloud_ICCV_2017_paper.html

[31] Yuan, M., Fu, K., Li, Z., Meng, Y., Wang, M.: PointMBF: A Multi-scale Bidirectional Fusion Network for Unsupervised RGB-D Point Cloud Registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 17694–17705 (2023). https://doi.org/10.1109/iccv51070.2023.01622

[32] Chen, C., Jia, X., Zheng, Y., Qu, Y.: RGBD-Glue: General Feature Combination for Robust RGB-D Point Cloud Registration. arXiv preprint arXiv:2405.07594 (2024). https://arxiv.org/abs/2405.07594

[33] Chen, C., Yan, S., Qu, Y.: ViGG: Robust RGB-D Point Cloud Registration using Visual-Geometric Mutual Guidance. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 816–826 (2026). https://doi.org/10.1109/wacv61042.2026.00086

[34] Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.-M.: USAC: A Universal Framework for Random Sample Consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8), 2022–2038 (2013). https://doi.org/10.1109/tpami.2012.257

[35] Barath, D., Matas, J.: Graph-Cut RANSAC. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6733–6741 (2018). https://doi.org/10.1109/cvpr.2018.00704

[36] Barath, D., Matas, J., Noskova, J.: MAGSAC: Marginalizing Sample Consensus. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10197–10205 (2019). https://openaccess.thecvf.com/content_CVPR_2019/html/Barath_MAGSAC_Marginalizing_Sample_Consensus_CVPR_2019_paper.html

[37] Barath, D., Noskova, J., Ivashechkin, M., Matas, J.: MAGSAC++, a Fast, Reliable and Accurate Robust Estimator. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1304–1312 (2020). https://doi.org/10.1109/cvpr42600.2020.00138

[38] Brachmann, E., Rother, C.: Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4322–4331 (2019). https://doi.org/10.1109/iccv.2019.00442

[39] Quan, S., Yang, J.: Compatibility-Guided Sampling Consensus for 3-D Point Cloud Registration. IEEE Transactions on Geoscience and Remote Sensing 58(10), 7380–7392 (2020). https://doi.org/10.1109/tgrs.2020.2982221

[40] Yang, J., Huang, Z., Quan, S., Qi, Z., Zhang, Y.: SAC-COT: Sample Consensus by Sampling Compatibility Triangles in Graphs for 3-D Point Cloud Registration. IEEE Transactions on Geoscience and Remote Sensing 60, 1–15 (2022). https://doi.org/10.1109/tg

rs.2021.3058552

[41] Belter, D., Nowicki, M., Skrzypczyński, P.: Modeling spatial uncertainty of point features in feature-based RGB-D SLAM. Machine Vision and Applications 29(5), 827–844 (2018). https://doi.org/10.1007/s00138-018-0936-9

[42] Choy, C., Dong, W., Koltun, V.: Deep Global Registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2514–2523 (2020). https://doi.org/10.1109/cvpr42600.2020.00259

[43] Bai, X., Luo, Z., Zhou, L., Chen, H., Li, L., Hu, Z., Fu, H., Tai, C.-L.: PointDSC: Robust Point Cloud Registration Using Deep Spatial Consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15859–15869 (2021). https://doi.org/10.1109/cvpr46437.2021.01560

[44] Arun, K.S., Huang, T.S., Blostein, S.D.: Least-Squares Fitting of Two 3-D Point Sets. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9(5), 698–700 (1987). https://doi.org/10.1109/tpami.1987.4767965

[45] Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 573–580 (2012). https://doi.org/10.1109/iros.2012.6385773

[46] Yang, J., Gao, Y., Li, D., Waslander, S.L.: ROBI: A Multi-View Dataset for Reflective Objects in Robotic Bin-Picking. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 9788–9795 (2021). https://doi.org/10.1109/iros51168.2021.9635871