Planar Shape Based Registration for Multi-modal Geometry

Muxingzi Li
muxingzi.li@inria.fr
Florent Lafarge
florent.lafarge@inria.fr

Inria, Université Côte d'Azur
Sophia Antipolis, France

Abstract

We present a global registration algorithm for multi-modal geometric data, typically 3D point clouds and meshes. Existing feature-based methods and recent deep learning based approaches typically rely upon point-to-point matching strategies that often fail to deliver accurate results from defect-laden data. In contrast, we reason at the scale of planar shapes whose detection from input data offers robustness on a range of defects, from noise to outliers through heterogeneous sampling. The detected planar shapes are projected into an accumulation space from which a rotational alignment is operated. A second step then refines the result with a local continuous optimization which also estimates the scale. We demonstrate the robustness and efficacy of our algorithm on challenging real-world data. In particular, we show that our algorithm competes well against state-of-the-art methods, especially on piece-wise planar objects and scenes.

1 Introduction

3D registration of multi-modal data is a long-standing challenge when working with real-world 3D objects. Geometric data obtained from different acquisition modalities (e.g. laser scans, multi-view stereo reconstruction) or created by modeling tools are represented in various forms, i.e. as point clouds or meshes, and exhibit different geometric properties in terms of noise, resolution or the scale. Classical problems in multi-modal registration involve registering a low-quality point cloud to a high-quality mesh, and registering a dense point cloud to a simplified mesh model.

Challenges in multi-modal registration arise from several aspects. Imperfection in data acquisition includes occlusions and non-uniform sampling density. Different surface representations, i.e. meshes and point clouds, often have different levels of detail and accuracy, making both traditional feature-based methods [32, 55, 57, 75] and deep learning architectures [3, 15, 40, 68, 69] unsuited for this task. Variation in acquisition modalities can lead to scale ambiguity, e.g. multi-view stereo generates data in an unknown scale, which further complicates the problem. The majority of existing methods [6, 12, 43, 49, 72, 74] focus on aligning 3D models to depth scans under the assumption that the model and the depth scan are already at the same scale. This is not the case for many real-world scenarios, where either the collected data or the 3D object model may have no absolute scale associated.

Simple pre-processing by estimating and correcting the scale before calling the registration step often fails for non-uniformly sampled data or partially overlapping data. Several works [17, 21, 28, 34, 44] have considered relative scale estimation. These methods treat the scale estimation as a separate step, therefore there lacks a unified formulation that simultaneously solve for the scale, rotation and translation.

In this work, we present a method for the global registration of multi-model geometric data of different scales, as illustrated in Fig. 1. It consists in, first, a rotational alignment that analyses the surface-normal distributions of the mesh and planar shapes detected from the input point set, and then a local refinement based on continuous optimization with Lie Algebra. The motivation behind our use of planar shapes arises from two aspects. First, planar shape detection methods, which have been successfully used in various vision tasks such as camera pose estimation [51], Structure from Motion [52, 77], or surface reconstruction [5], offer robustness to noise, outliers and varying sampling density, as opposed to directly working with raw point clouds. Second, it gives a natural approximation of the distance field of the underlying surface of the point cloud. The surface-normal representation is invariant to scaling and translation, which enables the estimation of the initial rotation matrix independently. In contrary to previous work, we formulate the scale estimation as a part of the continuous optimization problem based on distance field in the refinement step, with no need of an initial guess for the scale. Our non-feature-based approach is robust to variations of levels of details, noise and sampling density across different inputs, and is suitable for processing large point clouds.

2 Related Work

We distinguish four families of methods for registering rigid 3D objects.

Local registration with known scale. ICP [6] is the best-known algorithm for finding the SE(3) transformation between surfaces. Variants of ICP [7, 24, 53, 54, 59, 74] are proposed to address different issues, such as radius of convergence, computational efficiency, noise, partiality and sparsity. Probabilistic approaches like EM-ICP [27] and Gaussian Mixture Model based methods [20, 22, 25, 36] are introduced for robustness to noise and outliers. Another branch of work concerns direct matching of distance functions [11, 13, 48, 60], which is more accurate and robust than ICP given sufficient spatial resolution.

Global registration with known scale. A popular family of methods involve establishing feature correspondences [32, 55, 57]. Fast Global Registration (FGR) [75] improves the inlier ratio of the correspondence set effectively by simple tests without recomputation. 4PCS [2] and Super4PCS [43] effectively lower the complexity of RANSAC by exploring the motion space with co-planar 4-point quadrilateral matching. Another family of methods use a branch-and-bound (BnB) strategy to exhaustively explore the solution space for a good optimum, but suffer from slow convergence (Go-ICP [72], GOSMA [12]). Fast rotation search algorithm with a new bounding function for BnB has been introduced for acceleration [49]. Eckart et al. [21] propose a multi-scale point matching process using a hierarchy of Gaussian Mixtures. Many works explore the use of alternative shape embedding. One family of methods utilize the Fourier transform to decouple rotation and translation [8, 38], but is sensitive to the voxel resolution. The signed distance field, encoded in a discrete voxel grid, is a popular implicit representation for registering depth images [10, 47, 61].

Learning-based methods with known scale. Recent advances in deep learning lead to the development of several neural networks for point cloud registration. The models can be roughly categorized as non-iterative and iterative methods. Non-iterative models have a natural speed advantage. Deep Closest Point [68] utilizes a transformer network for feature matching coupled with SVD for point-to-point registration. DeepGMR [73] avoids point-to-point matches by integrating the network inside a probabilistic registration paradigm: this solution reduces complexity while improving robustness. Iterative methods are believed to be more robust to partially overlapping inputs [3, 15, 40, 69]. In particular, PointNetLK [3] adapts PointNet into the Lucas-Kanade algorithm. PRNet [69] extends DCP to an iterative pipeline with keypoint detection designed for partial-to-partial registration. IDAM [40] proposes a distance-aware similarity matrix convolution for finding correspondences. Deep Global Registration [15] is an end-to-end 6D ConvNet built upon FCGF [23] and works well on real-world dataset.

Multi-modal registration with an unknown scale. The registration of multi-modal geometric data often involves estimating the relative scale between different types of data, e.g. when aligning a CAD model to a point cloud scan, and when registering volumetric images obtained from different modalities. A survey covering issues and methods related to this task can be found in [56]. The most straightforward method which simply normalizes scales in pre-processing [33] is unsuitable for partially overlapping and noisy data. Extensions of ICP integrate scale factor estimation by including a separate minimization step [78], by incorporating a bounded scale matrix [19], by registering cumulative contribution rate curves [41], or by using the maximum correntropy criterion [71]. Coherent Point Drift [45] and its extensions [30, 31] formulate the task as a probability density estimation problem and re-parametrizes GMM centroids with rigid parameters including the scale. Corsini et al. [17] extend 4PCS and propose a method for point-cloud-to-3D-model registration. Bulow et al. [9] extend the Fourier transform approach to incorporate scale estimation. Paudel et al. [50] formulate the task as a point-to-plane assignment problem utilizing a plane-based assumption of the 3D scene. Mellado et al. [44] introduce a descriptor based on Growing Least Squares for scale-invariant matching. Registration of 3D images from different scan modalities is an important task in medical imaging [42], where level-set algorithms [18, 64, 67] are widely applied. Another sub-family of methods concern aligning CAD models from a collection of pre-specified categories to depth scans. These approaches determine the scale via object detection in terms of 3D bounding boxes, but are limited to training categories. Among these studies, Song et al. [62] assume that the gravity direction is known and estimate rotation only around the gravity axis. Gupta et al. [28] rely on traditional ICP for aligning the input point set and the point set rendered from the model. Izadinia et al. [35] proposes a learning-based ICP approach which formulates the rotation estimation problem as a policy learning task for viewpoint prediction. Deformation of the CAD model is considered by a few works [4, 34, 46] for better fitting. Our approach differs from the above pipelines by integrating scale, rotation and translation into a single optimization framework.

3 Algorithm

We consider as input a pair of 3D data composed of a point cloud and a surface mesh which we denote by the source and the target respectively. The relative scale between them is un-

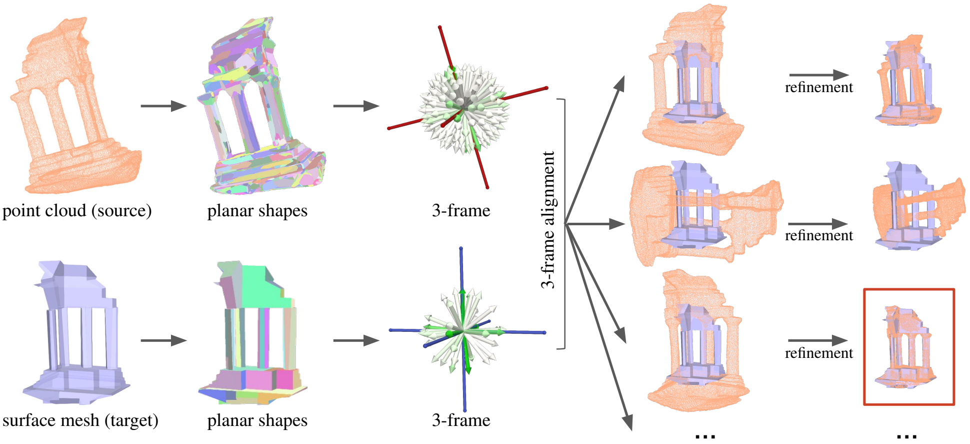


Figure 1: Overview of the proposed method. Planar shapes are first extracted from the input point cloud and surface mesh. From the surface-normal distribution of planar shapes (green corresponds to a high portion of planar area with normal pointing towards the arrow direction), three dominant directions are estimated, called a 3-frame. The 3-frames are aligned between the source and the target, leading to 24 possible rotations (only three are represented here). The refinement step takes each candidate rotation and estimate a final similarity transform. The alignment with the minimal loss is kept as the final result (see red frame).

known and the overlap can be partial. The goal is to determine the parameters of a similarity transformation S which best aligns the source against the target,

$$S = \begin{bmatrix} sR & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \tag{1}$$

where $s \in \mathbb{R}$, $R \in \mathbb{R}^{3\times 3}$ and $\mathbf{t} \in \mathbb{R}^3$ are the scale factor, the rotation matrix and the translation vector respectively.

The application of distance function representation removes the need for explicitly solving for correspondences. At first glance, it is intuitive to formulate the task as a least squares problem in the same way as rigid registration [24] using the distance field. Let $\{\mathbf{d}_i\}_1^{n_d}$ be a set of n_d points from the source, and $D_m : \mathbf{p} \in \mathbb{R}^3 \mapsto d \in \mathbb{R}$ be the distance field of the target surface, which maps a 3D point $\mathbf{p}$ to its Euclidean distance d to the closest point on the surface. Simple adaptation of the rigid registration formulation leads to a loss function given by

$$U(S) = \sum_{i=1}^{n_d} |D_m(S\mathbf{d}_i)|^2 \tag{2}$$

where conversion from homogeneous coordinates to Cartesian coordinates is omitted for simplicity of notations. The above formulation, however, has an infinite number of global minima $U = 0$ at scale factor $s = 0$, where the source simply shrinks to a single point on the target. These undesirable global minima result from the difference between Euclidean transformation and similarity transformation.

We propose an improved formulation by considering also the distance field D_d of the underlying surface of the source. Let $\{\mathbf{m}_i\}_1^{n_m}$ denote a set of n_m points sampled from the target. The proposed loss is given by

$$U(S) = \frac{1}{n_d}\sum_{i=1}^{n_d} |D_m(S\mathbf{d}_i)|^2 + \frac{1}{n_m}\sum_{j=1}^{n_m} |D_d(S^{-1}\mathbf{m}_j)|^2 \tag{3}$$

where S^{-1} is the inverse of S. The new loss is a symmetric measure of fit between the source and the target, which eliminates undesirable global minima. The distance fields are normalized beforehand. In order to minimize the proposed loss, we propose a two-step pipeline which consists in a rotational alignment followed by a local refinement. Fig. 1 shows an overview of our method.

3.1 Planar shape based alignment

The first step of our method consists in aligning the orientations of the source and target in a simple yet effective manner. The method generates a set of candidate rotation matrices, which will be refined in the later refinement step. First, a set of planar shapes are detected on the point cloud via region growing [39], with fixed parameters for all experiments. This step helps filtering out noisy points in the point cloud and yields an as clean as possible representation of the actual shapes, similar to the idea of Corsini et al. [17] who uses Variational Shape Approximation [16] to partition the point cloud into planar regions. From now on, we will discard the original point cloud and use the clean subset instead.

We propose to initialize the rotation matrix by aligning surface normals of the planar shapes of the source and the target. The alignment of normal vectors is invariant to scale and translation, which offers a more robust estimation. Our approach shares similarity with Stata Center World (SCW) [66] and the Manhattan Frame [63], which analyze the surface-normal distributions of a single input. In our setup, we focus on the relationship between the surface-normal distributions of the source and target. As shown in Fig. 1, we first cluster the normal vectors of each set of planar shapes to find 3 major axes (not necessarily orthogonal), which from now on will be called a 3-frame. A 3-frame represents the component means of a weighted-data Gaussian mixture model. In case of the point cloud, the data points are the alpha-shapes of the detected planes, weighted by their areas. In case of the surface mesh, they are the polygonal facets of the mesh, weighted by their areas. The distance metric of data points is defined on the unit sphere, where each axis includes both positive and negative directions. More specifically, a 3-frame can be represented as columns in

$$A = \begin{bmatrix}\mathbf{u}_1 & -\mathbf{u}_1 & \mathbf{u}_2 & -\mathbf{u}_2 & \mathbf{u}_3 & -\mathbf{u}_3\end{bmatrix}, \ \mathbf{u}_i \in \mathbb{R}^3. \tag{4}$$

We use the absolute cosine similarity metric instead of the Euclidean distance for measuring the distance between two normal vectors on the spherical surface, i.e. $d(\mathbf{v}_1, \mathbf{v}_2) = \frac{|\mathbf{v}_1 \cdot \mathbf{v}_2|}{\|\mathbf{v}_1\|\|\mathbf{v}_2\|}$.

We use the weighted Expectation-Maximization (EM) algorithm [26] to solve for the cluster means and variances. Same is done to the normal vectors of the target surface. The proposed approach is based on the assumption that, for the same underlying scene, there exists some column permutation P such that the 3-frames of different representations are aligned via a rotation R,

$$A_2 = RP(A_1). \tag{5}$$

For a given permutation P_i, the rotation matrix is computed as the solution to the orthogonal

Procrutes problem

$$R_i = \text{argmin}_{\Omega} \|\Omega P_i(A_1) - A_2\|_F \\ \text{subject to } \ \Omega^T \Omega = I \text{ and } \det(\Omega) = 1,$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The solution for R is obtained by only allowing orthogonal matrices with determinant 1, and is given by $R_i = U\Sigma' V^T$, where $A_2 P_i(A_1)^T = U\Sigma V^T$ is the singular value decomposition and Σ' is a modified Σ with the smallest singular value replaced by $\det(UV^T)$, and other singular values replaced by 1. The output of the rotational alignment step is a list of 24 rotation matrices $\{R_i\}_{i=1}^{24}$, taking into account all possible permutations with a consistent orientation of axes, where the consistent orientation means both following the right-hand rule or the left-hand rule.

3.2 Refinement

The results of the rotational alignment are now refined using local continuous optimization. For the input point cloud, the detected alpha shapes from the previous step are used to generate its distance field D_d. In order to solve the minimization problem locally, we rewrite the transformation matrix as $S = \exp(\boldsymbol{\xi})$ where $\boldsymbol{\xi} \in \mathbb{R}^7$ is the corresponding element in Lie algebra. We denote $\exp(\boldsymbol{\xi})$ as $S_{\boldsymbol{\xi}}$ from now on, and let $s_{\boldsymbol{\xi}}$ be the associated scale factor. The optimization objective becomes

$$\min_{\boldsymbol{\xi}} U(\boldsymbol{\xi}) = \frac{1}{n_d} \sum_{i=1}^{n_d} |D_m(S_{\boldsymbol{\xi}} \mathbf{d}_i)|^2 + \frac{1}{n_m} \sum_{j=1}^{n_m} |D_d(S_{\boldsymbol{\xi}}^{-1} \mathbf{m}_j)|^2. \tag{6}$$

The derivative of the loss is thus

$$\frac{dU}{d\boldsymbol{\xi}} = \frac{2}{n_d} \sum_{i=1}^{n_d} D_m(S_{\boldsymbol{\xi}} \mathbf{d}_i) \nabla D_m(S_{\boldsymbol{\xi}} \mathbf{d}_i) \frac{dS_{\boldsymbol{\xi}} \mathbf{d}_i}{d\boldsymbol{\xi}} + \frac{2}{n_m} \sum_{j=1}^{n_m} D_d(S_{\boldsymbol{\xi}}^{-1} \mathbf{m}_j) \nabla D_d(S_{\boldsymbol{\xi}}^{-1} \mathbf{m}_j) \frac{dS_{\boldsymbol{\xi}}^{-1} \mathbf{m}_j}{d\boldsymbol{\xi}} \tag{7}$$

where $\nabla D_m(\mathbf{p})$ is the gradient vector of the distance field at point $\mathbf{p}$. We have, for any point,

$$\frac{dS_{\boldsymbol{\xi}} \mathbf{p}}{d\boldsymbol{\xi}} = \begin{bmatrix} I & -\mathbf{q}'^{\wedge} & \mathbf{q}' \\ \mathbf{0}^T & \mathbf{0}^T & 0 \end{bmatrix} \in \mathbb{R}^{4\times 7}, \tag{8}$$

$$\frac{dS_{\boldsymbol{\xi}}^{-1} \mathbf{p}}{d\boldsymbol{\xi}} = \frac{dS_{-\boldsymbol{\xi}} \mathbf{p}}{d\boldsymbol{\xi}} \tag{9}$$

where $\mathbf{q}'$ denotes the Cartesian representation of the homogeneous coordinates $S_{\boldsymbol{\xi}} \mathbf{p}$, and $\mathbf{q}'^{\wedge}$ is the skew-symmetric matrix associated with vector $\mathbf{q}'$. Note that $S_{-\boldsymbol{\xi}} = \exp(-\boldsymbol{\xi}) = S_{\boldsymbol{\xi}}^{-1}$. Trust region methods, such as Levenberg-Marquardt and Dogleg, can be used for optimization. In our experiments, we use the Dogleg algorithm.

In our implementation, note that AABB tree is used for fast distance queries against sets of plane objects. Each query returns a closest point $\mathbf{p}$ on the set of planes to the query point $\mathbf{q}$. It also allows efficient computation of distance field gradient, as the gradient of a distance field always has magnitude 1 and has the same direction as $\mathbf{q} - \mathbf{p}$ whenever only one closest point exists.

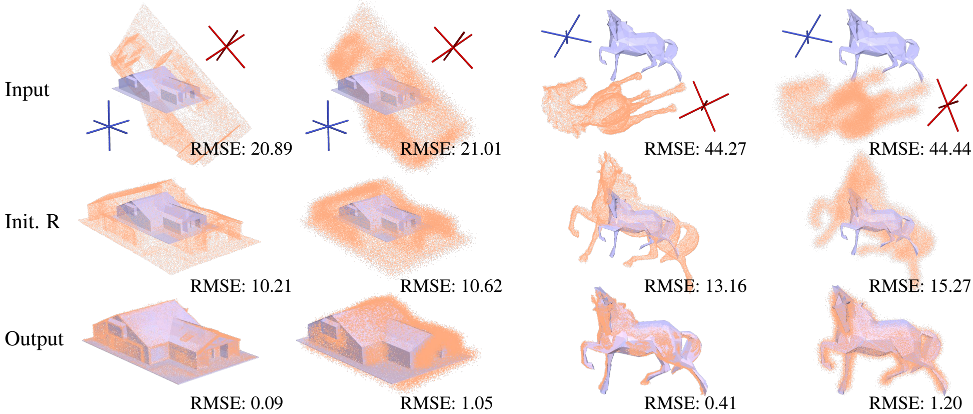


Figure 2: Visualization of our registration results on a regular object (barn) and a free-form object (horse). For each object, Gaussian noise is added to create a noisy version. The top row shows the mesh and the point cloud to be aligned, as well as their 3-frames (drawn in blue and red, respectively). The best initial rotation is shown in the middle row, with the aligned result at bottom. The RMSE ($\times 10^{-2}$) values are shown.

4 Experiments

Our algorithm is implemented in C++ using the Computational Geometry Algorithms Library (CGAL) [65] and the Ceres library [1]. For all experiments, the parameters of the region growing step for shape detection are fixed and set as follows: The Euclidean distance threshold is set to 4μ, where μ is the mean of K nearest neighbor distances of the point cloud. The normal threshold is set to 35 degrees. The minimum number of points per planar shape is 40.

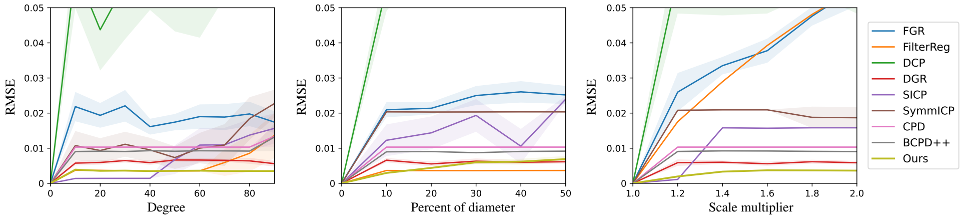
Dataset and error metrics. Existing datasets for rigid registration consist of point cloud pairs obtained from the same acquisition modality, which does not offer differences in terms of levels of detail and defects. Also, there is often no associated mesh data of the captured scene. Synthetic range data from meshes do not simulate well defects of real-world acquisition systems. To this end, we evaluate and compare our approach with state-of-the-art methods on a collection of 13 real-world point sets that differ in terms of shape complexity, size, and acquisition characteristics, provided in [5], together with their corresponding simplified 3D models. The point sets are acquired via different modalities, i.e. multi-view stereo, and laser scanner. Additionally, we include synthetic data consisting of point sets sampled from 6 shapes, and their simplified models computed using [5]. The simplified 3D models are compact mesh representations of the objects, which differ from the point sets in terms of detail, noise and outliers. The models are set to different scales in the experiments. The dataset is divided into two categories: free-form objects and regular objects. We consider an object to be regular if it exhibits a high degree of organization in the form of large planar structures, e.g. buildings and furniture. The rest are considered as free-form objects.

We use two metrics for quantitative evaluation: root mean square error (RMSE) and α-recall, similar to [75]. We compute the RMSE between the estimated transformation

$S = (s, R, t)$ from the ground-truth transformation $S^* = (s^*, R^*, t^*)$:

$$\varepsilon = \sqrt{\frac{1}{n_d} \sum_{i=1}^{n_d} \min_j \|sR\mathbf{d}_i + t - s^* R^* \mathbf{d}_j - t^*\|_2^2} \tag{10}$$

The RMSE is computed in a slightly different way from [75]: because we have no ground-truth pointwise correspondences in our point-to-mesh alignment setting, the distance is measured against the nearest neighbor in the ground-truth. Unlike some previous works, we do not directly compare against the ground-truth transformation nor on the distance between ground-truth correspondences in order to eliminate ambiguity for shapes with rotational symmetry, where multiple ground truths exist. α-recall is defined as the ratio of successful pairwise registrations, where a registration is considered successful if its RMSE is smaller than a certain threshold α. For both metrics, the RMSE unit is the diameter of the target.



(a) Rotation perturbation (b) Translation perturbation (c) Scale perturbation

Figure 3: The mean (bold curve) and standard deviation (shaded region) of the RMSE of each method on different perturbations to the ground-truth alignment. Lower is better.

Robustness. Fig. 2 shows our registration results for regular and free-form objects. To demonstrate robustness to noise, 3D Gaussian noise is added to the MVS point cloud of each object (2nd and 4th columns). Comparing the 3-frames of point clouds with and without added noise, it can be seen that estimation of 3-frames is robust to noise, especially for the case of regular objects. Thus our method is able to generate good initial rotations, which are refined to recover the final alignment.

We also investigate robustness to each type of perturbation (rotation, translation and scale). The results are shown in Fig. 3. In the first two experiments, the source is perturbed from the ground-truth alignment with varying degrees of rotation (or translation, resp.), keeping the ground-truth translation (or rotation, resp.) and scale. In the third experiment, the source is resized by a specific amount each time and undergoes randomly generated small rotations and translations. As illustrated in Fig. 3, our algorithm show competitive stability to all three types of perturbations.

Comparisons. We compare with both local and global methods. The local methods include SICP [78], SymmICP [64], CPD [45], BCPD++ [31] and FilterReg [25], while the global methods are FGR [75], DCP [68], and DGR [15]. In our experiment, all local methods are combined with a RANSAC initialization. Some methods have a built-in scale estimation component and are labeled as *joint* in Table 1. The others assume a given scale as input and are labeled as *two-step*. For these two-step methods, we provide an estimated scale factor from a bounding box based estimation method which can provide a good estimate given sufficient overlap between input [17, 44]. For all data tested in Table 1, the error is around

2.5% from the ground truth scale. Details of the scale estimation method can be found in supplementary material.

Table 1: Quantitative comparisons. Average (and maximal) RMSE ($\times 10^{-2}$) is computed over 50 random perturbations on scaling (between 0.25 and 4), rotation (between 60° and 180°), and translation (between 0 and 100% of the diameter) for each of the 19 models.

		free-form objects									regular objects									
		capron	horse	ignatius	m60	dragon	bunny	hand	rocker	eight	cottage	chair	bldgA	room	block	temple	barn	euler	hilbert	dice
two-step	FGR[[illegible]]	3.532 (9.471)	3.509 (8.427)	3.829 (9.266)	4.173 (7.949)	1.280 (2.415)	6.992 (8.472)	5.358 (10.280)	5.169 (12.768)	11.762 (17.103)	4.999 (8.307)	6.482 (8.502)	8.437 (13.041)	5.099 (8.194)	7.873 (15.813)	4.012 (11.484)	3.414 (7.434)	3.213 (6.776)	3.443 (4.832)	1.435 (4.261)
	FilterReg[[illegible]]	1.680 (2.576)	3.250 (5.232)	2.568 (5.297)	1.732 (2.873)	2.696 (5.374)	3.591 (6.153)	2.861 (4.583)	2.568 (4.245)	0.720 (0.735)	1.695 (2.907)	3.627 (6.125)	2.625 (4.348)	2.007 (3.286)	42.352 (>100)	3.770 (9.126)	1.832 (3.236)	1.566 (2.068)	0.853 (3.228)	1.048 (1.537)
	DCP[[illegible]]	6.439 (9.791)	9.028 (13.373)	5.174 (10.540)	6.977 (9.148)	5.985 (10.837)	6.933 (9.761)	5.098 (7.834)	8.666 (11.323)	7.095 (16.750)	4.362 (6.875)	6.200 (7.757)	8.902 (13.061)	4.919 (8.367)	9.509 (15.283)	8.966 (14.689)	8.294 (12.157)	6.551 (10.537)	3.088 (4.149)	1.410 (1.561)
	DGR[[illegible]]	1.377 (2.129)	0.339 (0.449)	0.259 **(0.418)**	**0.718** **(1.084)**	0.247 (0.400)	0.498 **(0.569)**	0.265 **(0.488)**	1.107 (3.651)	1.414 (8.287)	1.738 (8.294)	0.770 (1.147)	**0.679** **(0.860)**	1.190 (4.120)	**0.370** **(0.511)**	0.372 **(0.506)**	0.154 **(0.255)**	0.529 (1.188)	3.025 (7.451)	0.758 (1.149)
joint	SICP[[illegible]]	**1.059** (1.904)	**0.096** **(0.122)**	1.644 (3.024)	1.231 (3.203)	0.045 **(0.063)**	1.937 (4.591)	1.837 (3.665)	2.075 (3.288)	1.674 (2.745)	0.495 (2.578)	2.322 (5.192)	2.583 (6.738)	1.723 (3.114)	1.409 (1.962)	1.560 (2.891)	0.696 (3.164)	1.024 (1.409)	1.345 (1.437)	0.786 (1.514)
	SymmICP[[illegible]]	2.344 (3.055)	3.699 (5.800)	2.618 (6.722)	2.053 (4.124)	3.176 (5.670)	4.889 (8.530)	3.374 (5.807)	3.097 (5.200)	0.421 (0.575)	2.022 (4.533)	4.685 (7.238)	3.582 (7.535)	2.714 (4.651)	1.755 (2.783)	5.152 (10.979)	3.519 (6.157)	1.887 (3.008)	0.864 (2.651)	0.475 (0.488)
	CPD[[illegible]]	1.723 (2.352)	2.670 (4.334)	2.615 (5.344)	3.186 (3.825)	2.299 (5.429)	3.224 (5.888)	2.970 (4.610)	2.180 (3.680)	0.405 (0.575)	1.967 (3.089)	4.981 (6.287)	3.031 (4.232)	2.223 (5.971)	1.349 (2.557)	3.055 (5.275)	1.740 (3.112)	1.829 (2.367)	1.221 (1.316)	1.100 (1.632)
	BCPD++[[illegible]]	1.568 (2.609)	2.862 (5.087)	2.581 (5.577)	1.522 (2.970)	2.196 (4.387)	2.851 (5.136)	2.728 (4.405)	2.169 (3.970)	0.435 (0.592)	2.086 (3.688)	3.201 (4.495)	2.595 (3.942)	2.039 (4.211)	1.544 (2.851)	3.128 (5.493)	1.807 (3.393)	1.710 (2.588)	2.363 (2.406)	21.311 (21.321)
	Ours	1.338 **(1.842)**	0.383 (5.035)	**0.136** (0.941)	1.720 (2.258)	**0.042** (1.332)	**0.314** (4.000)	**0.170** (2.144)	**0.095** **(1.300)**	**0.314** **(0.562)**	**0.427** **(1.211)**	**0.268** **(0.696)**	1.188 (5.417)	**0.348** **(1.047)**	0.920 (1.316)	**0.170** (0.850)	**0.079** (0.286)	**0.253** **(0.742)**	**0.039** **(0.513)**	**0.487** **(0.504)**

Our algorithm performs best on regular objects and scenes as they are well described by piecewise-planar geometry. As shown in the right part of Table 1, our method reaches significantly lower average RMSE in 8 out of 10 objects, while retaining errors reasonably close to the best baseline in the remaining 2 cases. In addition, our method achieves the lowest maximal RMSE for almost half of the tested objects, exhibiting a low failure rate on regular objects comparable to other methods. The failure case, bldgA, among regular objects is due to the noisy normal of input points. On the contrary, the best performing baseline, DGR, is less robust as its maximal RMSE tends to be off by a large amount when it fails, *e.g.* on cottage, hilbert and dice. Fig. 4 (b) shows the α-recall rate of all methods on all tests done on the free-form objects. Our algorithm achieves a 0.02-recall of 99%, significantly higher than the other algorithms, with DGR reaching 88%. For free-form objects, as indicated by the left part of Table 1 as well as Fig. 4 (a), our method matches the accuracy achieved by state-of-the-art methods. Visual comparisons are provided in supplementary material.

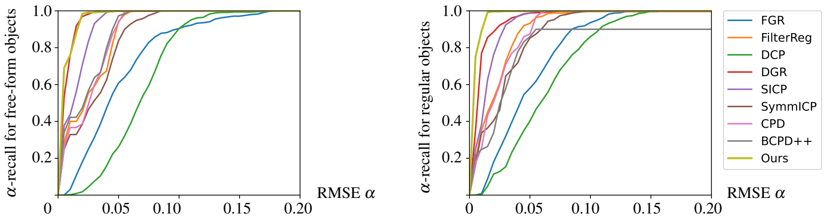


Figure 4: α-recall curve of each method on free-form objects (left) and regular objects (right). In particular, our approach outperforms existing methods on regular objects.

Experiment on multi-way registration. We also evaluate the efficacy of our approach on the multi-way registration task. Similarly to Choy et al.[15], we follow the multi-way registration pipeline proposed in [75], and replace the pairwise registration stage with our proposed method. We adapt our algorithm so that both source and target are point clouds. The experiment is performed on the Augmented ICL-NUIM dataset [14, 29]. Accuracy is measured as the absolute trajectory error (ATE) defined in [14].

Although our method is not originally designed to perform on this task where scans do not necessarily overlap well, it offers promising results compared to online SLAM algorithms and offline multi-way reconstruction methods. As shown in Table 2, our method obtains an higher accuracy than ElasticFusion [70], InfiniTAM [37], BAD-SLAM [58] and FGR [75] for almost all the scenes. Our accuracy remains slightly lower than the multi-way reconstruction methods [15, 76] which have been designed to perform on this task.

Table 2: Quantitative comparisons on multi-way registration from the Augmented ICL-NUIM dataset. The accuracy scores correspond to the ATE error expressed in centimeter, where lower is better.

	ElasticFusion [70]	InfiniTAM [37]	BAD-SLAM [58]	Multi-way + FGR [75]	Multi-way + RANSAC [76]	Multi-way + DGR [15]	Multi-way + Ours
Living room 1	66.61	46.07	fail	78.97	110.9	21.06	31.98
Living room 2	24.33	73.64	40.41	24.91	19.33	21.88	23.60
Office 1	13.04	113.8	18.53	14.96	14.42	15.76	19.68
Office 2	35.02	105.2	26.34	21.05	17.31	11.56	21.03

Performances. The planar shape-based alignment typically requires a few seconds to one minute depending on the size of the input point cloud (that ranges from 150K to 3M points). This corresponds to the processing time for detecting planar shapes, clustering being negligible. The refinement step is also a few seconds for each rotational initialization from our non-optimized sequential implementation of the algorithm.

Limitations. Our algorithm, which is designed to perform on regular scenes, is less competitive on free-form objects. The detection of planar shapes on such objects often gives a rough and arbitrary approximation of their curved surfaces. Our method is also not designed to the registration of 3D data with a very low overlap ratio.

5 Conclusion

We proposed a global registration algorithm for multi-modal geometric data which differs in terms of noise, detail, and scales. Our algorithm performs a planar shape based alignment to recover candidate rotations independent of scale and translation, followed by a refinement step with a local continuous optimization. We demonstrated the robustness and efficacy of our algorithm on defect-laden real-world data, as well as it competitiveness against state-of-the-art methods, especially on objects and scenes that can be described with a piece-wise planar geometry.

In future work, we plan to extend our method to partial-to-partial registration with very low overlap ratio between input geometry. One way could be to design a confidence estimation method for weighing each data point. The weight can be assigned according to the likelihood of having the point also contained in the other input. The estimated weight can be combined with both our rotational alignment step and refinement step.

Acknowledgments. This work was partially supported by ANR-17-CE23-0003 project BIOM.

References

[1] Sameer Agarwal, Keir Mierle, and Others. Ceres solver. http://ceres-solver.org.

[2] Dror Aiger, Niloy J. Mitra, and Daniel Cohen-Or. 4-points congruent sets for robust surface registration. *ACM Transactions on Graphics*, 27(3), 2008.

[3] Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. PointNetLK: Robust & efficient point cloud registration using PointNet. In *CVPR*, 2019.

[4] Armen Avetisyan, Manuel Dahnert, Angela Dai, Manolis Savva, Angel X. Chang, and Matthias Niessner. Scan2cad: Learning cad model alignment in rgb-d scans. In *CVPR*, 2019.

[5] Jean-Philippe Bauchet and Florent Lafarge. Kinetic shape reconstruction. *ACM Transactions on Graphics*, 39(5), 2020.

[6] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *TPAMI*, 14(2), 1992.

[7] Sofien Bouaziz, Andrea Tagliasacchi, and Mark Pauly. Sparse iterative closest point. In *Symposium on Geometry Processing*, 2013.

[8] Heiko Bülow and Andreas Birk. Spectral 6DOF registration of noisy 3d range data with partial overlap. *TPAMI*, 35(4), 2013.

[9] Heiko Bülow and Andreas Birk. Scale-free registrations in 3d: 7 degrees of freedom with fourier mellin soft transforms. *IJCV*, 126, 2018.

[10] E. Bylow, Jürgen Sturm, C. Kerl, F. Kahl, and D. Cremers. Real-time camera tracking and 3d reconstruction using signed distance functions. In *Robotics: Science and Systems*, 2013.

[11] Erik Bylow, Jurgen Sturm, Christian Kerl, and Fredrik Kahl. Real-time camera tracking and 3d reconstruction using signed distance functions. In *RSS*, 2013.

[12] Dylan Campbell, Lars Petersson, Laurent Kneip, Hongdong Li, and Stephen Gould. The alignment of the spheres: Globally-optimal spherical mixture alignment for camera pose estimation. In *CVPR*, 2019.

[13] Daniel R. Canelhas, Todor Stoyanov, and Achim J. Lilienthal. Sdf tracker: A parallel algorithm for on-line pose estimation and scene reconstruction from depth images. In *IROS*, 2013.

[14] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *CVPR*, 2015.

[15] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *CVPR*, 2020.

[16] David Cohen-Steiner, Pierre Alliez, and Mathieu Desbrun. Variational shape approximation. *ACM Transactions on Graphics*, 23(3), 2004.

[17] Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, Riccardo Gherardi, Andrea Fusiello, and Roberto Scopigno. Fully automatic registration of image sets on approximate geometry. *IJCV*, 102, 2013.

[18] D. Cremers, Mikaël Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *IJCV*, 72:195–215, 2006.

[19] Shaoyi Du, Nanning Zheng, Shihui Ying, Qubo You, and Yang Wu. An extension of the ICP algorithm considering scale factor. In *ICIP*, 2007.

[20] Ben Eckart, Kihwan Kim, Alejandro Troccoli, Alonzo Kelly, and Jan Kautz. MLMD: Maximum likelihood mixture decoupling for fast and accurate point cloud registration. In *International Conference on 3D Vision (3DV)*, 2015.

[21] Ben Eckart, Kihwan Kim, and Jan Kautz. Fast and accurate point cloud registration using trees of gaussian mixtures. In *ECCV*, 2018.

[22] Georgios D. Evangelidis and Radu Horaud. Joint alignment of multiple point sets with batch and incremental expectation-maximization. *TPAMI*, 40(6), 2018.

[23] Qiaojun Feng and Nikolay Atanasov. Fully convolutional geometric features for category-level object alignment. In *IROS*, 2020.

[24] Andrew Fitzgibbon. Robust registration of 2d and 3d point sets. In *BMVC*, 2001.

[25] Wei Gao and Russ Tedrake. Filterreg: Robust and efficient probabilistic point-set registration using gaussian filter and twist parameterization. In *CVPR*, 2019.

[26] Israel Dejene Gebru, Xavier Alameda-Pineda, Florence Forbes, and Radu Horaud. Em algorithms for weighted-data clustering with application to audio-visual scene analysis. *TPAMI*, 38(12), 2016.

[27] Sébastien Granger and Xavier Pennec. Multi-scale em-icp: A fast and robust approach for surface registration. In *ECCV*, 2002.

[28] Saurabh Gupta, Pablo AndrÃl's ArbelÃąez, Ross B. Girshick, and Jitendra Malik. Aligning 3d models to rgb-d images of cluttered scenes. In *CVPR*, 2015.

[29] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J. Davison. A benchmark for rgb-d visual odometry, 3d reconstruction and slam. In *ICRA*, 2014.

[30] Osamu Hirose. A bayesian formulation of coherent point drift. *TPAMI*, 43(7), 2021.

[31] Osamu Hirose. Acceleration of non-rigid point set registration with downsampling and gaussian process regression. *TPAMI*, 2021.

[32] Dirk Holz, Alexandru E. Ichim, Federico Tombari, Radu B. Rusu, and Sven Behnke. Registration with the point cloud library: A modular framework for aligning in 3-d. *IEEE Robotics Automation Magazine*, 22(4), 2015.

[33] Xiaoshui Huang, Jian Zhang, Lixin Fan, Qiang Wu, and Chun Yuan. A systematic approach for cross-source point cloud registration by preserving macro and micro structures. *IEEE Transactions on Image Processing*, 26(7), 2017.

[34] Vladislav Ishimtsev, Alexey Bokhovkin, Alexey Artemov, Savva Ignatyev, Matthias Niessner, Denis Zorin, and Evgeny Burnaev. Cad-deform: Deformable fitting of cad models to 3d scans. In *ECCV*, 2020.

[35] Hamid Izadinia and Steven M. Seitz. Scene recomposition by learning-based icp. In *CVPR*, 2020.

[36] Bing Jian and Baba C. Vemuri. Robust point set registration using gaussian mixture models. *TPAMI*, 33(8), 2011.

[37] Olaf Kähler, Victor Prisacariu, and David Murray. Real-time large-scale dense 3d reconstruction with loop closure. In *ECCV*, 2016.

[38] Yosi Keller, Yoel Shkolnisky, and Amir Averbuch. Volume registration using the 3-d pseudopolar fourier transform. *IEEE Trans. on Signal Processing*, 54(11), 2006.

[39] Florent Lafarge and Clement Mallet. Creating large-scale city models from 3d-point clouds: A robust approach with hybrid representation. *IJCV*, 99, 2012.

[40] Jiahao Li, Changhao Zhang, Ziyao Xu, Hangning Zhou, and Chi Zhang. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration. In *ECCV*, 2020.

[41] Baowei Lin, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, and Koji Ichii. Scale ratio ICP for 3d point clouds with different scales. In *ICIP*, 2013.

[42] J.B.Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1), 1998.

[43] Nicolas Mellado, Dror Aiger, and Niloy J. Mitra. Super 4PCS fast global pointcloud registration via smart indexing. *Computer Graphics Forum*, 33(5), 2014.

[44] Nicolas Mellado, Matteo Dellepiane, and Roberto Scopigno. Relative scale estimation and 3d registration of multi-modal geometry using growing least squares. *IEEE Transactions on Visualization and Computer Graphics*, 22(9), 2016.

[45] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *TPAMI*, 32(12), 2010.

[46] Liangliang Nan, Ke Xie, and Andrei Sharf. A search-classify approach for cluttered indoor scene understanding. *ACM Transactions on Graphics*, 31, 11 2012.

[47] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *IEEE International Symposium on Mixed and Augmented Reality*, 2011.

[48] Nikos Paragios, Mikael Rousson, and Visvanathan Ramesh. Matching distance functions: A shape-to-area variational approach for global-to-local registration. In *ECCV*, 2002.

[49] Álvaro Parra Bustos, Tat-Jun Chin, Anders Eriksson, Hongdong Li, and David Suter. Fast rotation search with stereographic projections for 3d registration. *TPAMI*, 38(11), 2016.

[50] Danda Pani Paudel, Adlane Habed, Cédric Demonceaux, and Pascal Vasseur. Robust and optimal sum-of-squares-based point-to-plane registration of image sets and structured scenes. In *ICCV*, 2015.

[51] Carolina Raposo, M. Lourenco, M. Antunes, and J. Barreto. Plane-based odometry using an rgb-d camera. In *BMVC*, 2013.

[52] Carolina Raposo, Michel Antunes, and Joao P. Barreto. Piecewise-planar stereoscan: Sequential structure and motion using plane primitives. *TPAMI*, 40(8), 2018.

[53] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *Proceedings of International Conference on 3-D Digital Imaging and Modeling (3DIM)*, 2001.

[54] Szymon Rusinkiewicz. A symmetric objective function for ICP. *ACM Transactions on Graphics*, 38(4), 2019.

[55] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *ICRA*, 2009.

[56] E. Saiti and T. Theoharis. An application independent review of multimodal 3d registration methods. *Computers & Graphics*, 91, 2020.

[57] Ruwen Schnabel, Roland Wahl, and Reinhard Klein. Efficient ransac for point-cloud shape detection. *Computer Graphics Forum*, 26(2), 2007.

[58] Thomas Schöps, Torsten Sattler, and Marc Pollefeys. Bad slam: Bundle adjusted direct rgb-d slam. In *CVPR*, 2019.

[59] Aleksandr Segal, Dirk Hahnel, and Sebastian Thrun. Generalized-icp. In *Proc. of Robotics: Science and Systems*, 2009.

[60] Miroslava Slavcheva, Wadim Kehl, Nassir Navab, and Slobodan Ilic. Sdf-2-sdf: Highly accurate 3d object reconstruction. In *ECCV*, 2016.

[61] Miroslava Slavcheva, Wadim Kehl, Nassir Navab, and Slobodan Ilic. Sdf-2-sdf registration for real-time 3d reconstruction from rgb-d data. *IJCV*, 126(6), 2018.

[62] Shuran Song and Jianxiong Xiao. Sliding shapes for 3d object detection in depth images. In *ECCV*, 2014.

[63] J. Straub, G. Rosman, O. Freifeld, J. J. Leonard, and J. W. Fisher. A mixture of manhattan frames: Beyond the manhattan world. In *CVPR*, 2014.

[64] Piotr Swierczynski, Bartlomiej W. Papiez, Julia A. Schnabel, and Colin Macdonald. A level-set approach to joint image segmentation and registration with application to ct lung imaging. *Computerized Medical Imaging and Graphics*, 65, 2018.

[65] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 5.2.1 edition, 2021. URL https://doc.cgal.org/5.2.1/Manual/packages.html.

[66] R. Triebel, W. Burgard, and F. Dellaert. Using hierarchical em to extract planes from 3d range scans. In *ICRA*, 2005.

[67] BC Vemuri, J Ye, Y Chen, and CM Leonard. Image registration via level-set motion: applications to atlas-based segmentation. *Medical image analysis*, 7(1), 2003.

[68] Yue Wang and Justin M. Solomon. Deep closest point: Learning representations for point cloud registration. In *ICCV*, 2019.

[69] Yue Wang and Justin M. Solomon. PRNet: Self-supervised learning for partial-to-partial registration. In *NIPS*, 2019.

[70] Thomas Whelan, Stefan Leutenegger, Renato F. Salas-Moreno, B. Glocker, and A. Davison. Elasticfusion: Dense slam without a pose graph. In *Robotics: Science and Systems*, 2015.

[71] Zongze Wu, Hongchen Chen, Shaoyi Du, Minyue Fu, Nan Zhou, and Nanning Zheng. Correntropy based scale icp algorithm for robust point set registration. *Pattern Recognition*, 93, 2019.

[72] Jiaolong Yang, Hongdong Li, and Yunde Jia. Go-ICP: Solving 3d registration efficiently and globally optimally. In *CVPR*, 2013.

[73] Wentao Yuan, Benjamin Eckart, Kihwan Kim, Varun Jampani, Dieter Fox, and Jan Kautz. DeepGMR: Learning latent gaussian mixture models for registration. In *ECCV*, 2020.

[74] Juyong Zhang, Yuxin Yao, and Bailin Deng. Fast and robust iterative closest point. *TPAMI*, 2021.

[75] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In *ECCV*, 2016.

[76] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018.

[77] Zihan Zhou, Hailin Jin, and Yi Ma. Robust plane-based structure from motion. In *CVPR*, 2012.

[78] Timo Zinßer, Jochen Schmidt, and Heinrich Niemann. Point set registration with integrated scale estimation. In *International Conference on Pattern Recognition and Image Processing*, 2005.