Density Adaptive Point Set Registration

Felix Järemo Lawin, Martin Danelljan, Fahad Shahbaz Khan, Per-Erik Forssén, Michael Felsberg
Computer Vision Laboratory, Department of Electrical Engineering, Linköping University, Sweden
{felix.jaremo-lawin, martin.danelljan, fahad.khan, per-erik.forssen, michael.felsberg}@liu.se

Abstract

Probabilistic methods for point set registration have demonstrated competitive results in recent years. These techniques estimate a probability distribution model of the point clouds. While such a representation has shown promise, it is highly sensitive to variations in the density of 3D points. This fundamental problem is primarily caused by changes in the sensor location across point sets. We revisit the foundations of the probabilistic registration paradigm. Contrary to previous works, we model the underlying structure of the scene as a latent probability distribution, and thereby induce invariance to point set density changes. Both the probabilistic model of the scene and the registration parameters are inferred by minimizing the Kullback-Leibler divergence in an Expectation Maximization based framework. Our density-adaptive registration successfully handles severe density variations commonly encountered in terrestrial Lidar applications. We perform extensive experiments on several challenging real-world Lidar datasets. The results demonstrate that our approach outperforms state-of-the-art probabilistic methods for multi-view registration, without the need of re-sampling.

1. Introduction

3D-point set registration is a fundamental problem in computer vision, with applications in 3D mapping and scene understanding. Generally, the point sets are acquired using a 3D sensor, *e.g.* a Lidar or an RGBD camera. The task is then to align point sets acquired at different positions, by estimating their relative transformations. Recently, probabilistic registration methods have shown competitive performance in different scenarios, including pairwise [19, 14, 15] and multi-view registration [10, 6].

In this work, we revisit the foundations of the probabilistic registration paradigm, leading to a reformulation of the Expectation Maximization (EM) based approaches [10, 6]. In these approaches, a Maximum Likelihood (ML) formulation is used to simultaneously infer the transformation parameters, and a Gaussian mixture model (GMM) of the point distribution. Our formulation instead minimizes the Kullback-Leibler divergence between the mixture model and a latent scene distribution.

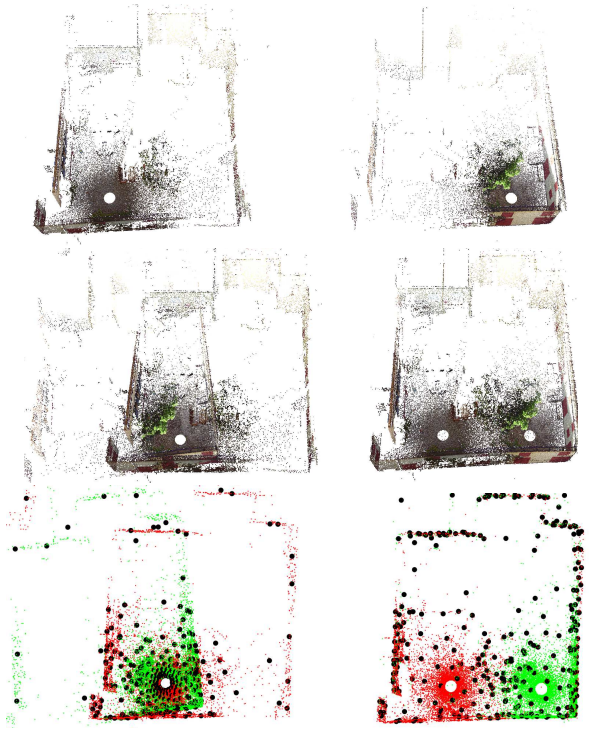


Figure 1. Two example Lidar scans (top row), with significantly varying density of 3D-points. State-of-the-art probabilistic method [6] (middle left) only aligns the regions with high density. This is caused by the emphasis on dense regions, as visualized by the Gaussian components in the model (black circles in bottom left). Our method (right) successfully exploits essential information available in sparse regions, resulting in accurate registration.

Common acquisition sensors, including Lidar and RGBD cameras, do not sample all surfaces in the scene with a uniform density (figure 1, top row). The density of 3D-point observations is highly dependent on (1) the distance to the sensor, (2) the direction of the surface relative to the sensor, and (3) inherent surface properties, such as specularity. Despite recent advances, state-of-the art prob-

abilistic methods [19, 10, 6, 15, 14] struggle under varying sampling densities, in particular when the translational part of the transformation is significant. The density variation is problematic for standard ML-based approaches since each 3D-point corresponds to an observation with equal weight. Thus, the registration focuses on regions with high point densities, while neglecting sparse regions.

This negligence is clearly visible in figure 1 (bottom left), where registration has been done using CPPSR [6]. Here the vast majority of Gaussian components (black circles) are located in regions with high point densities. A common consequence of this is inaccurate or failed registrations. Figure 1 (middle right) shows an example registration using our approach. Unlike the existing method [6], our model exploits information available in both dense and sparse regions of the scene, as shown by the distribution of Gaussian components (figure 1, bottom right).

1.1. Contributions

We propose a probabilistic point set registration approach that counters the issues induced by sampling density variations. Our approach directly models the underlying structure of the 3D scene using a novel density-adaptive formulation. The probabilistic scene model and the transformation parameters are jointly inferred by minimizing the Kullback-Leibler (KL) divergence with respect to the latent scene distribution. This is enabled by modeling the acquisition process itself, explicitly taking the density variations into account. To this end, we investigate two alternative strategies for estimating the acquisition density: a model-based and a direct empirical method. Experiments are performed on several challenging Lidar datasets, demonstrating the effectiveness of our approach in difficult scenarios with drastic variations in the sampling density.

2. Related work

The problem of 3D-point set registration is extensively pursued in computer vision. Registration methods can be coarsely categorized into local and global methods. Local methods rely on an initial estimate of the relative transformation, which is then iteratively refined. The typical example of a local method is the Iterative Closest Point (ICP) algorithm. In ICP, registration is performed by iteratively alternating between establishing point correspondences and refining the relative transformation. While the standard ICP [1] benefits from a low computational cost, it is limited by a narrow region of convergence. Several works [23, 21, 4] investigate how to improve the robustness of ICP.

Global methods instead aim at finding the global solution to the registration problem. Many global methods rely on local ICP-based or probabilistic methods and use, *e.g.*, multiple restarts [17], graph optimization [24], branch-and-bound [3] techniques to search for a globally optimal registration. Another line of research is to use feature descriptors to find point correspondences in a robust estimation framework, such as RANSAC [20]. Zhou *et al*. [27] also use feature correspondences, but minimize a Geman-McClure robust loss. A drawback of such global methods is the reliance on accurate geometric feature extraction.

Probabilistic registration methods model the distribution of points as a density function. These methods perform alignment either by employing a correlation based approach or using an EM based optimization framework. In correlation based approaches [25, 15], the point sets are first modeled separately as density functions. The relative transformation between the points set is then obtained by minimizing a metric or divergence between the densities. These methods lead to nonlinear optimization problems with non-convex constraints. Unlike correlation based methods, the EM based approaches [19, 10] find an ML-estimate of the density model and transformation parameters.

Most methods implicitly assume a uniform density of the point clouds, which is hardly the case in most applications. The standard approach [22] to alleviate the problems of varying point density is to re-sample the point clouds in a separate preprocessing step. The aim of this strategy is to achieve an approximately uniform distribution of 3D points in the scene. A common method is to construct a voxel grid and taking the mean point in each voxel. Comparable uniformity is achieved using the Farthest Point Strategy [8], were points are selected iteratively to maximize the distance to neighbors. Geometrically Stable Sampling (GSS) [11] also incorporates surface normals in the sample selection process. However, such re-sampling methods have several shortcomings. First, 3D scene information is discarded as observations are grouped together or removed, leading to sparsification of the point cloud. Second, the sampling rate, *e.g.* voxel size, needs to be hand picked for each scenario as it depends on the geometry and scale of the point cloud. Third, a suitable trade-off between uniformity and sparsity must be found. Thus, such preprocessing steps are complicated and their efficacy is questionable. In this paper, we instead explicitly model the density variations induced by the sensor.

There exist probabilistic registration methods that tackle the problem of non-uniform sampling density [2, 13]. In [2], a one class support vector machine is trained for predicting the underlying density of partly occluded point sets. The point sets are then registered by minimizing the L2 distance between the density models. In [16], an extended EM framework for modeling noisy data points is derived, based on minimizing the KL divergence. This framework was later exploited for outlier handling in point set registration [13]. Unlike these methods, we introduce a latent distribution of the scene and explicitly model the point sampling density using either a sensor model or an empirical method.

3. Method

In this work, we revisit probabilistic point cloud registration, with the aim of alleviating the problem of non-uniform point density. To show the impact of our model, we employ the Joint Registration of Multiple Point Clouds (JRMPC) [10]. Compared to previous probabilistic methods, JRMPC has the advantage of enabling joint registration of multiple input point clouds. Furthermore, this framework was recently extended to use color [6], geometric feature descriptors [5] and incremental joint registration [9]. However, our approach can be applied to a variety of other probabilistic registration approaches. Next, we present an overview of the baseline JRMPC method.

3.1. Probabilistic Point Set Registration

Point set registration is the problem of finding the relative geometric transformations between M different sets of points. We directly consider the general case where $M \geq 2$. Each set $\mathcal{X}_i = \{x_{ij}\}_{j=1}^{N_i}, i = 1, \dots, M$, consists of 3D-point observations $x_{ij} \in \mathbb{R}^3$ obtained from, e.g., a Lidar scanner or an RGBD camera. We let capital letters X_{ij} denote the associated random variables for each observation. In general, probabilistic methods aim to model the probability densities $p_{X_i}(x)$, for each point set i, using for instance Gaussian Mixture Models (GMMs).

Different from previous approaches, JRMPC derives the densities $p_{X_i}(x)$ from a global probability density model $p_V(v|\theta)$, which is defined in a reference coordinate frame given parameters θ. The registration problem can then be formulated as finding the relative transformations from point set $\mathcal{X}_i$ to the reference frame. We let $\phi(\cdot;\omega) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be a 3D transformation parametrized by $\omega \in \mathbb{R}^D$. The goal is then to find the parameters ω_i of the transformation from $\mathcal{X}_i$ to the reference frame, such that $\phi(X_{ij};\omega_i) \sim p_V$. Similarly to previous works [10, 6], we focus on the most common case of rigid transformation $\phi(x;\omega) = R_\omega x + t_\omega$. In this case, the density model of each point set is obtained as $p_{X_i}(x|\omega_i,\theta) = p_V(\phi(x;\omega_i)|\theta)$.

The density $p_V(v|\theta)$ is composed by a mixture of Gaussian distributions,

$$p_V(v|\theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(v;\mu_k,\Sigma_k)\,. \tag{1}$$

Here, $\mathcal{N}(v;\mu,\Sigma)$ is a Gaussian density with mean μ and covariance Σ. The number of components is denoted by K and π_k is the prior weight of component k. The set of all mixture parameters is thus $\theta = \{\pi_k,\mu_k,\Sigma_k\}_{k=1}^K$.

Different from previous works, the mixture model parameters θ and transformation parameters ω are inferred jointly in the JRMPC framework, assuming independent observations. This is achieved by maximizing the log-likelihood function,

$$\mathcal{L}(\Theta;\mathcal{X}_1,\dots,\mathcal{X}_M) = \sum_i^M \sum_j^{N_i} \log(p_V(\phi(x_{ij};\omega_i)|\theta))\,. \tag{2}$$

Here, we denote the set of all parameters in the model as $\Theta = \{\theta,\omega_1,\dots,\omega_M\}$. Inference is performed with the Expectation Maximization (EM) algorithm, by first introducing a latent variable $Z \in \{1,\dots,K\}$ that assigns a 3D-point V to a particular mixture component $Z = k$. The complete data likelihood is then given by $p_{V,Z}(v,k|\theta) = p_Z(k|\theta)p_{V|Z}(v|k,\theta)$, where $p_Z(k|\theta) = \pi_k$ and $p_{V|Z}(v|k,\theta) = \mathcal{N}(v;\mu_k,\Sigma_k)$. The original mixture model (1) is recovered by marginalizing the complete data likelihood over the latent variable Z.

The E-step in the EM algorithm involves computing the expected complete-data log likelihood,

$$Q(\Theta;\Theta^n) = \sum_i^M \sum_j^{N_i} E_{Z|x_{ij},\Theta^n}[\log(p_{V,Z}(\phi(x_{ij};\omega_i),Z|\theta))]\,. \tag{3}$$

Here, the conditional expectation is taken over the latent variable given the observed point x_{ij} and the current estimate of the model parameters Θ^n. In the M-step, the model parameters are updated as $\Theta^{n+1} = \arg\max_\Theta Q(\Theta;\Theta^n)$. This process is then repeated until convergence.

3.2. Sampling Density Adaptive Model

To tackle the issues caused by non-uniform point densities, we revise the underlying formulation and model assumptions. Instead of modeling the density of 3D-points, we aim to infer a model of the actual 3D-structure of the scene. To this end, we introduce the latent probability distribution of the scene $q_V(v)$. Loosely defined, it is seen as a uniform distribution on the observed surfaces in the scene. Intuitively, $q_V(v)$ encodes all 3D-structure, i.e. walls, ground, objects etc., that is measured by the sensor. Different models of $q_V(v)$ are discussed is section 3.4. Technically, q_V might not be absolutely continuous and is thus regarded a probability measure. However, we will denote it as a density function to simplify the presentation.

Our goal is to model $q_V(v)$ as a parametrized density function $p_V(v|\theta)$. We employ a GMM (1) and minimize the Kullback-Leibler (KL) divergence from p_V to q_V,

$$\mathrm{KL}(q_V||p_V) = \int \log\left(\frac{q_V(v)}{p_V(v|\theta)}\right) q_V(v)\,\mathrm{d}v\,. \tag{4}$$

Utilizing the decomposition of the KL-divergence $\mathrm{KL}(q_V||p_V) = H(q_V,p_V) - H(q_V)$ into the cross entropy $H(q_V,p_V)$ and entropy $H(q_V)$ of q_V, we can equivalently maximize,

$$\mathcal{E}(\Theta) = -H(q_V,p_V) = \int \log\left(p_V(v|\theta)\right) q_V(v)\mathrm{d}v \tag{5}$$

In (5), the integration is performed in the reference frame of the scene. On the other hand, the 3D points x_{ij} are observed in the coordinate frames of the individual sensors. As in section 3.1, we relate these coordinate frames with the transformations $\phi(\cdot;\omega_i)$. By applying the change of variables $v = \phi(x;\omega_i)$, we obtain

$$\mathcal{E}(\Theta) = \frac{1}{M}\sum_{i=1}^{M}\int_{\mathbb{R}^3} \log\left(p_V(\phi(x;\omega_i)|\theta)\right)\cdot \qquad (6)$$
$$q_V(\phi(x;\omega_i))|\det(D\phi(x;\omega_i))|\,\mathrm{d}x\,.$$

Here, $|\det(D\phi(x;\omega_i))|$ is the determinant of the Jacobian of the transformation. From now on, we assume rigid transformations, which implies $|\det(D\phi(x;\omega_i))| = 1$.

We note that if $\{x_{ij}\}_{i=1}^{N_i}$ are independent samples from $q_V(\phi(x;\omega_i))$, the original maximum likelihood formulation (2) is recovered as a Monte Carlo sampling of the objective (6). Therefore, the conventional ML formulation (2) relies on the assumption that the observed points x_{ij} follow the underlying uniform distribution of the scene q_V. However, this assumption completely neglects the effects of the acquisition sensor. Next, we address this problem by explicitly modeling the sampling process.

In our formulation, we consider the points in set i to be independent samples $x_{ij} \sim q_{X_i}$ of a distribution $q_{X_i}(x)$. In addition to the 3D structure q_V of the scene, q_{X_i} can also depend on the position and properties of the sensor, and the inherent properties of the observed surfaces. This enables more realistic models of the sampling process to be employed. By assuming that the distribution q_V is absolutely continuous [7] w.r.t. q_{X_i}, eq. (6) can be written,

$$\mathcal{E}(\Theta) = \sum_{i=1}^{M}\int_{\mathbb{R}^3} \log\left(p_V(\phi(x;\omega_i)|\theta)\right)\frac{q_V(\phi(x;\omega_i))}{q_{X_i}(x)}q_{X_i}(x)\,\mathrm{d}x. \qquad (7)$$

Here, we have also ignored the factor $1/M$. The fraction $f_i(x) = \frac{q_V(\phi(x;\omega_i))}{q_{X_i}(x)}$ is known as the Radon-Nikodym derivative [7] of the probability distribution $q_V(\phi(x;\omega_i))$ with respect to $q_{X_i}(x)$. Intuitively, $f_i(x)$ is the ratio between the density in the latent scene distribution and the density of points in point cloud $\mathcal{X}_i$. Since it weights the observed 3D-points based on the local density, we term it the *observation weighting function*. In section 3.4, we later introduce two different approximations of $f_i(x)$ to model the sampling process itself.

3.3. Inference

In this section, we describe the inference algorithm used to minimize (7). We show that the EM-based framework used in [10, 6] also generalizes to our model. As in section 3.1, we apply the latent variable Z and the complete-data likelihood $p_{V,Z}(v,k|\theta)$. We define the expected complete-data cross entropy as,

$$\mathcal{Q}(\Theta,\Theta^n) = \qquad (8)$$
$$\sum_{i=1}^{M}\int_{\mathbb{R}^3} E_{Z|x,\Theta^n}\left[\log\left(p_{V,Z}(\phi(x;\omega_i),Z|\theta)\right)\right] f_i(x) q_{X_i}(x)\,\mathrm{d}x.$$

Here, Θ^n is the current estimate of the parameters. The E-step involves evaluating the expectation in (8), taken over the probability distribution of the latent variable,

$$p_{Z|X_i}(k|x,\Theta) = \frac{p_{X_i,Z}(x,k|\Theta)}{\sum_{k=1}^{K} p_{X_i,Z}(x,k|\Theta)} = \frac{\pi_k\mathcal{N}(\phi(x;\omega_k);\mu_k,\Sigma_k)}{\sum_{l=1}^{K}\pi_l\mathcal{N}(\phi(x;\omega_l);\mu_l,\Sigma_l)}\,. \qquad (9)$$

To maximize (8) in the M-step, we first perform a Monte Carlo sampling of (8). Here we use the assumption that the observations are independent samples drawn from $x_{ij} \sim q_{X_i}$. To simplify notation, we define $\alpha_{ijk}^n = p_{Z|X_i}(k|x_{ij},\Theta^n)$. Then (8) is approximated as,

$$\mathcal{Q}(\Theta,\Theta^n) \approx Q(\Theta,\Theta^n) = \qquad (10)$$
$$\sum_{i=1}^{M}\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{K}\alpha_{ijk}^n f_i(x_{ij})\log\left(p_{V,Z}(\phi(x_{ij};\omega_i),k|\theta)\right)\,.$$

Please refer to the supplementary material for a detailed derivation of the EM procedure.

The key difference of (10) compared to the ML case (3), is the weight factor $f_i(x_{ij})$. This factor effectively weights each observation x_{ij} based on the local density of 3D points. Since the M-step has a form similar to (3), we can apply the optimization procedure proposed in [10]. Specifically, we employ two conditional maximization steps [18], to optimize over the mixture parameters θ and transformation parameters ω_i respectively. Furthermore, our approach can be extended to incorporate color information using the approach proposed in [6].

3.4. Observation Weights

We present two approaches of modeling the observation weight function $f_i(x)$. The first is based on a sensor model, while the second is an empirical estimation of the density.

3.4.1 Sensor Model Based

Here, we estimate the sampling distribution q_{X_i} by modeling the acquisition sensor itself. For this method we therefore assume that the type of sensor (e.g. Lidar) is known and that each point set $\mathcal{X}_i$ consists of a single scan. The latent scene distribution q_V is modeled as a uniform distribution on the observed surfaces S. That is, S is a 2-dimensional manifold consisting of all observable surfaces.

Thus, we define $q_V(A) = \frac{1}{|S|}\int_{S\cap A} \mathrm{d}S$ for any measurable set $A \subset \mathbb{R}^3$. For simplicity, we use the same notation $q_V(A) = \mathbb{P}(V \in A)$ for the probability measure q_V of V. We use $|S| = \int_S \mathrm{d}S$ to denote the total area of S.

We model the sampling distribution q_{X_i} based on the properties of a terrestrial Lidar. It can however be extended to other sensor geometries, such as time-of-flight cameras. We can without loss of generality assume that the Lidar is positioned in the origin $x = 0$ of the sensor-based reference frame in $\mathcal{X}_i$. Further, let $S_i = \phi_i^{-1}(S)$ be the scene transformed to the reference frame of the sensor. Here, we use $\phi_i(x) = \phi(x, \omega_i)$ to simplify notation. We note that the density of Lidar rays is decreasing quadratically with distance. For this purpose, we model the Lidar as light source emitting uniformly in all directions of its field of view. The sampling probability density at a visible point $x \in S_i$ is then proportional to the absorbed intensity, calculated as $\frac{\hat{n}_x^{\mathrm{T}}\hat{x}}{\|x\|^2}$. Here, $\hat{n}_x$ is the unit normal vector of S_i at x, $\|\cdot\|$ is the Euclidean norm and $\hat{x} = x/\|x\|$.

The sampling distribution is defined as the probability of observing a point in a subset $A \subset \mathbb{R}^3$. It is obtained by integrating the point density over the part of the surface S intersecting A,

$$q_{X_i}(A) = \int_{S_i\cap A} \frac{g_i}{|S|}\,\mathrm{d}S_i, \quad g_i(x) = \begin{cases} a\frac{\hat{n}_x^{\mathrm{T}}\hat{x}}{\|x\|^2}, x \in S_i \cap F_i \\ \varepsilon \qquad\quad , \text{otherwise} \end{cases} \tag{11}$$

Here, $F_i \subset \mathbb{R}^3$ is the observed subset of the scene, ε is the outlier density and a is a constant such that the probability integrates to 1. Using the properties of q_V, we can rewrite (11) as $q_{X_i}(A) = \int_A g_i\,\mathrm{d}(q_V \circ \phi_i)$. Here, $q_V \circ \phi_i$ is the composed measure $q_V(\phi_i(A))$. From the properties of the Radon-Nikodym derivative [7], we obtain that $f_i = \frac{\mathrm{d}(q_V\circ\phi_i)}{\mathrm{d}q_{X_i}} = \frac{1}{g_i}$. In practice, surface normal estimates can be noisy, thus promoting the use of a regularized quotient $f_i(x) = a\frac{\|x\|^2}{\gamma\hat{n}_x^{\mathrm{T}}\hat{x}+1-\gamma}$, for some fix parameter $\gamma \in [0, 1]$. Note that the calculation of $f_i(x)$ only requires information about the distance $\|x\|$ to the sensor and the normal $\hat{n}_x$ of the point cloud at x. For details and derivations, see the supplementary material.

3.4.2 Empirical Sample Model

As an alternative approach, we propose an empirical model of the sampling density. Unlike the sensor-based model in section 3.4.1, our empirical approach does not require any information about the sensor. It can thus be applied to arbitrary point clouds, without any prior knowledge. We modify the latent scene model q_V from sec. 3.4.1 to include a 1-dimensional Gaussian distribution in the normal direction of the surface S. This uncertainty in the normal direction models the coarseness or evenness of the surface, which leads to

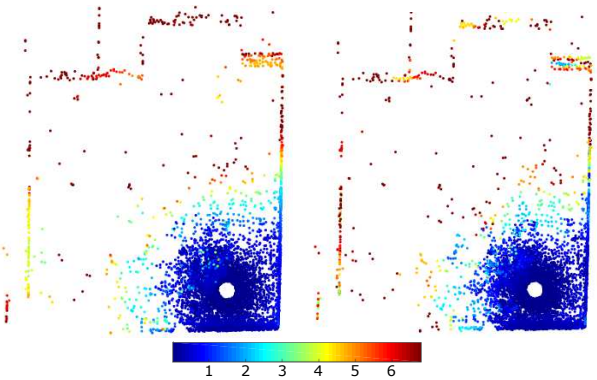


Figure 2. Visualization of the observation weight computed using our sensor based model (left) and empirical method (right). The 3D-points in the densely sampled regions in the vicinity of the Lidar are assigned low weights, while the impact of points in the sparser regions are increased. The two approaches produce visually similar results. The main differences are seen in the transitions from dense to sparser regions.

variations orthogonal to the underlying surface. In the local neighborhood of a point $\bar{v} \in S$, we can then approximate the latent scene distribution as a 1-dimensional Gaussian in the normal direction $q_V(v) \approx \frac{1}{|S|}\mathcal{N}(\hat{n}_{\bar{v}}^{\mathrm{T}}(v - \bar{v}); 0, \sigma_{\hat{n}}^2(\bar{v}))$. It is motivated by a locally planar approximation of the surface S at $\bar{v}$, where $q_V(v)$ is constant in the tangent directions of S. Here, $\sigma_{\hat{n}}^2(\bar{v})$ is the variance in the normal direction.

To estimate the observation weight function $f(x) = \frac{q_V(\phi(x))}{q_X(x)}$, we also find a local approximation of the sampling density $q_X(x)$. For simplicity, we drop the point set index i in this section and assume a rigid transformation $\phi(x) = Rx + t$. First, we extract the L nearest neighbors $x_1, \ldots, x_L$ of the 3D point x in the point cloud. We then find the local mean $\bar{x} = \frac{1}{L}\sum_l x_l$ and covariance $C = \frac{1}{L-1}\sum_l (x_l - \bar{x})^{\mathrm{T}}(x_l - \bar{x})$. This yields the local sampling density estimate $q_X(x) \approx \frac{L}{N}\mathcal{N}(x; \bar{x}, C)$. Let $C = BDB^{\mathrm{T}}$ be the eigenvalue decomposition of C with $B = (\hat{b}_1, \hat{b}_2, \hat{b}_3)$ and $D = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, and eigenvalues sorted in descending order. Since we assume the points to originate from a locally planar region, we deduce that $\sigma_1^2, \sigma_2^2 \gg \sigma_3^2$. Furthermore, $\hat{b}_3$ and σ_3^2 approximate the normal direction of the surface and the variance in this direction. We utilize this information for estimating the local latent scene distribution, by setting $\bar{v} = \phi(\bar{x})$, $\hat{n}_{\bar{v}} = R\hat{b}_3$ and $\sigma_{\hat{n}}^2(\bar{v}) = \sigma_3^2$. We then obtain,

$$f(x) = \frac{q_V(\phi(x))}{q_X(x)} \propto \sigma_1\sigma_2 e^{\frac{1}{2}(x-\bar{x})^{\mathrm{T}}B\begin{pmatrix}\sigma_1^{-2} & 0 & 0\\ 0 & \sigma_2^{-2} & 0\\ 0 & 0 & 0\end{pmatrix}B^{\mathrm{T}}(x-\bar{x})}. \tag{12}$$

Here, we have omitted proportionality constants independent of the point location x in $f(x)$, since they do not influence the objective (7). A detailed derivation is provided in the supplementary material. In practice, we found $f(x) \propto \sigma_1\sigma_2$ to be a sufficiently good approximation since

$\sigma_1^{-2}, \sigma_2^{-2} \approx 0$ and $\bar{x} \approx x$.

Note that the observation weights $f_i(x_{ij})$ in (10) can be precomputed once for every registration. The added computational cost of the density adaptive registration method is therefore minimal and in our experiments we only observed an increase in computational time of 2% compared to JRMPC. In figure 2, the observation weights $f_i(x_{ij})$ are visualized for both the sensor based model (left) and empirical method (right).

4. Experiments

We integrate our sampling density adaptive model in the probabilistic framework JRMPC [10]. Furthermore, we evaluate our approach, when using feature information, by integrating the model in the color based probabilistic method CPPSR [6].

First we perform a synthetic experiment to highlight the impact of sampling density variations on point set registration. Second, we perform quantitative and qualitative evaluations on two challenging Lidar scan datasets: Virtual Photo Sets [26] and the ETH TLS [24]. Further detailed results are presented in the supplementary material.

4.1. Experimental Details

Throughout the experiments we randomly generate ground-truth rotations and translations for all point sets. The point sets are initially transformed using this ground-truth. The resulting point sets are then used as input for all compared registration methods. For efficiency reasons we construct a random subset of 10k points for each scan in all the datasets. The experiments on the point sets from VPS and ETH TLS are conducted in two settings. First, we perform direct registration on the constructed point sets. Second, we evaluate all compared registration methods, except for our density adaptive model, on re-sampled point sets. The registration methods without density adaptation, however, are sensitive to the choice of re-sampling technique and sampling rate. In the supplementary material we provide an exhaustive evaluation of FPS [8], GSS [11] and voxel grid re-sampling at different sampling rates. We then extract the best performing re-sampling settings for each registration method and use it in the comparison as an empirical upper bound in performance.

Method naming: We evaluate two main variants of the density adaptive model. In the subsequent performance plots and tables, we denote our approach using the sensor model based observation weights in section 3.4.1 by DARS, and the empirical observation weights in section 3.4.2 by DARE.

Parameter settings: We use the same values for all the parameters that are shared between our methods and the two baselines: the JRMPC and CPPSR. As in [10], we use a uniform mixture component to model the outliers.

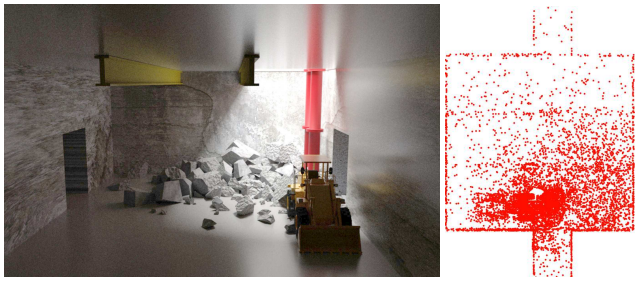


Figure 3. The synthetic 3D scene. Left: Rendering of the scene. Right: Top view of re-sampled point set with varying density.

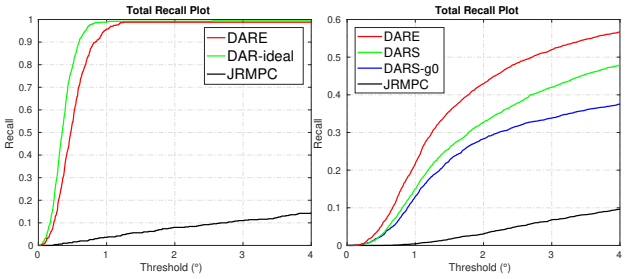
In our experiments, we set the outlier ratio 0.005 and fix the spatial component weights π_k to uniform. In case of pairwise registration, we set the number of spatial components $K = 200$. In the joint registration scenario, we set $K = 300$ for all methods to increase the capacity of the model for larger scenes. We use 50 EM iterations for both the pairwise and joint registration scenarios. In case of color features, we use 64 components as proposed in [6].

In addition to the above mentioned parameters, we use the $L = 10$ nearest neighbors to estimate σ_1 and σ_2 in section 3.4.2. To regularize the observation weights $f_i(x_{ij})$ (section 3.4) and remove outlier values, we first perform a median filtering using the same neighborhood size of $L = 10$ points. We then clip all the observation weights that exceed a certain threshold. We fix this threshold to 8 times the mean value of all observation weights within a point set. In the supplementary material we provide an analysis of these parameters and found our method not to be sensitive to the parameter values. For the sensor model approach (section 3.4.1) we set $\gamma = 0.9$. We keep all parameters fix in all experiments and datasets.

Evaluation Criteria: The evaluation is performed by computing the angular error (*i.e.* the *geodesic distance*) between the found rotation, R, and the ground-truth rotation, R_{gt}. This distance is computed via the Frobenius distance $d_F(R, R_{\text{gt}})$, using the relation $d_G(R_1, R_2) = 2\sin^{-1}(d_F(R_1, R_2)/\sqrt{8})$, which is derived in [12]. To evaluate the performance in terms of robustness, we report the failure rate as the percentage of registrations with an angular error greater than 4 degrees. Further, we present the accuracy in terms of the mean angular error among inlier registrations. In the supplementary material we also provide the translation error.

4.2. Synthetic Data

We first validate our approach on a synthetic dataset to isolate the impact of sampling density variations on pairwise registration. We construct synthetic point clouds by performing point sampling on a polygon mesh that simulates an indoor 3D scene (see figure 3 left). We first sample uniformly, and densely. We then randomly select a virtual



(a) Synthetic (b) Combined VPS and TLS

Figure 4. Recall curves with respect to the angular error. (a) Results on the synthetic dataset. Our DARE approach closely follows the upper bound, DAR-ideal. (b) Results on the combined VPS and TLS ETH datasets. In all cases, our DARE approach significantly improves over the baseline JRMPC [10].

	Avg. inlier error (°)	Failure rate (%)
JRMPC	2.44±0.87	90.4
JRMPC-*eub*	1.67±0.92	46.0
ICP	1.73±1.04	62.6
ICP-*eub*	1.81±0.99	55.7
CPD	1.88±1.25	90.0
CPD-*eub*	1.30±0.95	40.8
DARE	1.45±0.89	43.3

Table 1. A comparison of our approach with existing methods in terms of average inlier angular error and failure rate for pairwise registration on the combined VPS and TLS ETH dataset. The methods with the additional *-eub* in the name are the empirical upper bounds using re-sampling. Our DARE method improves over the baseline JRMPC, regardless of re-sampling settings, both in terms of accuracy and robustness.

sensor location. Finally, we simulate Lidar sampling density variations by randomly removing points according to their distances to the sensor position (see figure 3 right). In total the synthetic dataset contains 500 point set pairs.

Figure 4a shows the recall curves, plotting the ratio of registrations with an angular error smaller than a threshold. We report results for the baseline JRMPC and our DARE method. We also report the results when using the ideal sensor sample model to compute the observation weights $f_i(x_{ij})$, called DAR-ideal. Note that the same sampling function was employed in the construction of the virtual scans. This method therefore corresponds to an upper performance bound of our DARE approach.

The baseline JRMPC model struggles in the presence of sampling density variations, providing inferior registration results with a failure rate of 85 %. Note that the JRMPC corresponds to setting the observation weights to uniform $f_i(x_{ij}) = 1$ in our approach. The proposed DARE, significantly improves the registration results by reducing the failure rate from 85 % to 2 %. Further, the registration performance of DARE closely follows the ideal sampling density model, demonstrating the ability of our approach to adapt to sampling density variations.

4.3. Pairwise Registration

We perform pairwise registration experiments on the joint Virtual Photo Set (VPS) [26] and the TLS ETH [24] datasets. The VPS dataset consists of Lidar scans from two separate scenes, each containing four scans. The TLS ETH dataset consists of two separate scenes, with seven and five scans respectively. We randomly select pairs of different scans within each scene, resulting in total 3720 point set pairs. The ground-truth for each pair is generated by first randomly selecting a rotation axis. We then rotate one of the point sets with a rotation angle (within 0-90 degrees) around the rotation axis and apply a random translation, drawn from a multivariate normal distribution with standard deviation 1.0 meters in all directions.

Table 4b shows pairwise registration comparisons in terms of angular error on the joint dataset. We compare the baseline JRMPC [10] with both of our sampling density models: DARE and DARS. We also show the results for DARS without using normals, *i.e.* setting $\gamma = 0$ in section 3.4.1, in the DARS-g0 curve. All the three variants of our density adaptive approach significantly improve over the baseline JRMPC [10]. Further, our DARE model provides the best results. It significantly reduces the failure rate from 90.4% to 43.3%, compared to the JRMPC method.

We also compare our empirical density adaptive model with several existing methods in the literature. Table 1 shows the comparison of our approach with the JRMPC [10], ICP[1] [1], and CPD [19] methods. We present numerical values for the methods in terms of average inlier angular error and the failure rate.

Additionally, we evaluate the existing methods using re-sampling. In the supplementary material we provide an evaluation of different re-sampling approaches at different sampling rates. For each of the methods JRMPC [10], ICP [1], and CPD [19], we select the best performing re-sampling approach and sampling rate. In practical applications however, such comprehensive exploration of the re-sampling parameters is not feasible. In this experiment, the selected re-sampling settings serve as empirical upper bounds, denoted by *-eub* in the method names in table 1.

From table 1 we conclude that regardless of re-sampling approach, our DARE still outperforms JRMPC, both in terms of robustness and accuracy. The best performing method overall was the empirical upper bound for CPD with re-sampling. However, CPD is specifically designed for pairwise registration, while JRMPC and our approach also generalize to multi-view registration.

4.4. Multi-view registration

We evaluate our approach in a multi-view setting, by jointly registering all four point sets in the VPS indoor

[1] We use the built-in Matlab implementation of ICP.

(a) CPPSR (b) DARE-color

Figure 5. Joint registration of the four point sets in the VPS indoor dataset. (a) CPPSR [6] only aligns the high density regions and neglects sparsely sampled 3D-structure. (b) Corresponding registration using our density adaptive model incorporating color information.

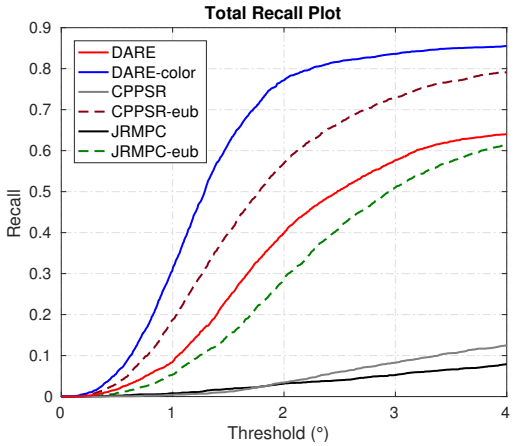


Figure 6. A multi-view registration comparison of our density adaptive model and existing methods, in terms of angular error on the VPS indoor dataset. Our model provides lower failure rate compared to the baseline methods JRMPC and CPPSR, also in comparison to the empirical upper bound.

	Avg. inlier error (°)	Failure rate (%)
CPPSR	2.57±0.837	87.4
CPPSR-*eub*	1.63±0.807	20.9
JRMPC	2.38±1.01	92.1
JRMPC-*eub*	2.13±0.83	38.6
DARE-color	1.26±0.61	14.5
DARE	1.84±0.80	36.0

Table 2. A multi-view registration comparison of our density adaptive model with existing methods in terms of average inlier angular error and failure rate on the VPS indoor dataset. Methods with -*eub* in the name are empirical upper bounds. Our model provides improved results, both in terms of robustness and accuracy.

dataset. We follow a similar protocol as in the pairwise registration case (see supplementary material). In addition to the JRMPC, we also compare our color extension with the CPPSR approach of [6]. Table 2 and figure 6 shows the multi-view registration results on the VPS indoor dataset. As in the pairwise scenario, the selected re-sampled versions are denoted by *-eub* in the method name. We use the same re-sampling settings for JRMPC and CPPSR as for JRMPC in the pairwise case. Both JRMPC and CPPSR have a significantly lower accuracy and a higher failure rate compared to our sampling density adaptive models. We further observe that re-sampling improves both JRMPC and CPPSR, however, not to the same extent as our density adaptive approach. Figure 5 shows a qualitative comparison between our color based approach and the CPPSR method [6]. In agreement with the pairwise scenario (see figure 1) CPPSR locks on to the high density regions, while our density adaptive approach successfully registers all scans, producing an accurate reconstruction of the scene. Further, we provide additional results on the VPS outdoor dataset in the supplementary material.

5. Conclusions

We investigate the problem of sampling density variations in probabilistic point set registration. Unlike previous works, we model both the underlying structure of the 3D scene and the acquisition process to obtain robustness to density variations. Further, we jointly infer the scene model and the transformation parameters by minimizing the KL divergence in an EM based framework. Experiments are performed on several challenging Lidar datasets. Our proposed approach successfully handles severe density variations commonly encountered in real-world applications.
Acknowledgements: This work was supported by the EU's Horizon 2020 Programme grant No 644839 (CENTAURO), CENIIT grant (18.14), and the VR grants: EMC2 (2014-6227), starting grant (2016-05543), LCMM (2014-5928).

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *PAMI*, 14(2):239–256, 1992. 2, 7

[2] D. Campbell and L. Petersson. An adaptive data representation for robust point-set registration and merging. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4292–4300, 2015. 2

[3] D. Campbell and L. Petersson. Gogma: Globally-optimal gaussian mixture alignment. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 2

[4] D. Chetverikov, D. Stepanov, and P. Krsek. Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm. *IMAVIS*, 23(3):299–309, 2005. 2

[5] M. Danelljan, G. Meneghetti, F. Shahbaz Khan, and M. Felsberg. Aligning the dissimilar: A probabilistic method for feature-based point set registration. In *ICPR*, 2016. 3

[6] M. Danelljan, G. Meneghetti, F. Shahbaz Khan, and M. Felsberg. A probabilistic framework for color-based point set registration. In *CVPR*, 2016. 1, 2, 3, 4, 6, 8

[7] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010. 4, 5

[8] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi. The farthest point strategy for progressive image sampling. *TIP*, 6(9):1305–1315, 1997. 2, 6

[9] G. D. Evangelidis and R. Horaud. Joint alignment of multiple point sets with batch and incremental expectation-maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. In press. Early Access. 3

[10] G. D. Evangelidis, D. Kounades-Bastian, R. Horaud, and E. Z. Psarakis. A generative model for the joint registration of multiple point sets. In *European Conference on Computer Vision*, pages 109–122. Springer, 2014. 1, 2, 3, 4, 6, 7

[11] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy. Geometrically stable sampling for the icp algorithm. In *3DIM 2003*, 2003. 2, 6

[12] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, July 2013. 6

[13] J. Hermans, D. Smeets, D. Vandermeulen, and P. Suetens. Robust point set registration using em-icp with information-theoretically optimal outlier handling. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2465–2472. IEEE, 2011. 2

[14] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang. Rigid and articulated point registration with expectation conditional maximization. *PAMI*, 33(3):587–602, 2011. 1, 2

[15] B. Jian and B. C. Vemuri. Robust point set registration using gaussian mixture models. *PAMI*, 33(8):1633–1645, 2011. 1, 2

[16] L. J. Latecki, M. Sobel, and R. Lakaemper. New em derived from kullback-leibler divergence. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 267–276. ACM, 2006. 2

[17] J. P. Luck, C. Q. Little, and W. Hoff. Registration of range data using a hybrid simulated annealing and iterative closest point algorithm. In *Proceedings of the 2000 IEEE International Conference on Robotics and Automation, ICRA 2000, April 24-28, 2000, San Francisco, CA, USA*, pages 3739–3744, 2000. 2

[18] X.-L. Meng and D. B. Rubin. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, 80(2):268–278, 1993. 4

[19] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010. 1, 2, 7

[20] R. Raguram, J.-M. Frahm, and M. Pollefeys. A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. *Computer Vision–ECCV 2008*, pages 500–513, 2008. 2

[21] A. Rangarajan, H. Chui, and F. L. Bookstein. The softassign procrustes matching algorithm. In *IPMI*, 1997. 2

[22] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *Robotics and automation (ICRA), 2011 IEEE International Conference on*, pages 1–4. IEEE, 2011. 2

[23] A. Segal, D. Hähnel, and S. Thrun. Generalized-icp. In *RSS*, 2009. 2

[24] P. W. Theiler, J. D. Wegner, and K. Schindler. Globally consistent registration of terrestrial laser scans via graph optimization. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109:126–138, 2015. 2, 6, 7

[25] Y. Tsin and T. Kanade. A correlation-based approach to robust point set registration. In *European conference on computer vision*, pages 558–569. Springer, 2004. 2

[26] J. Unger, A. Gardner, P. Larsson, and F. Banterle. Capturing reality for computer graphics applications. In *Siggraph Asia Course*, 2015. 6, 7

[27] Q.-Y. Zhou, J. Park, and V. Koltun. Fast global registration. In *European Conference on Computer Vision*, pages 766–782. Springer, 2016. 2