Scan Registration using the Normal Distributions
Transform with Region Growing Clustering for
Point-Sampled 3D Surfaces

Siddhant Ahuja* and Steven L. Waslander†
University of Waterloo, Waterloo, ON, N2L 3G1, Canada

The normal distributions transform (NDT) registration algorithm often converges to a local minimum if there is a large initial transformation error. In order to improve convergence properties, a novel region growing clustering NDT (RGC-NDT) algorithm is proposed, which replaces volumetric divisions and removes discontinuities in the cost function at the voxel boundaries during the optimization step. A set of computationally efficient multi-scale difference of locally salient feature vectors is computed for each point in the point cloud using principal components analysis (PCA). Adjacent points around the seed point in the local neighborhood are aggregated into clusters according to the region membership criterion based on the similarity of these feature vectors. The normal distributions transform is then computed for each cluster and points within the region are represented as a probability density function (PDF) for distribution-to-distribution matching. Scan registration results from indoor laser scans validate the improvement of the basin of convergence for NDT algorithm over existing methods.

I. Introduction

3D mapping of unstructured environments relies on accurate alignment of partially overlapping scans into a globally consistent model, called scan registration. Sensors such as RGB-D cameras, LIDAR, Time of Flight (ToF), and stereo cameras provide information as point-sampled 3D surfaces, termed point-clouds. Overlapping scans share a common set of points that can be used for matching in order to estimate the relative rigid body transformation between scans (6-DOF rotation and translation). Separate views of the same environment can be accumulated into a global coordinate system which helps an intelligent mobile robot perform tasks in an unstructured environment. However, points within each scan represent samples of different surfaces within the environment, subject to the type of sensor used for capturing the scene, sampling density (number of points per volumetric unit), sensor viewpoint (relative geometric position), sensitivity to measurement noise, quantization errors, occlusions, depth-discontinuities due to sharp edges, and the surface characteristics of the objects within the scene such as color, shape, textures, etc. (see Figure 1). Finding accurate transformation parameters, given the intra-scan problems outlined earlier and a relatively large initial inter-scan transformation error, makes the registration problem especially hard.

One of the most popular scan registration methods-iterative closest point (ICP) [2–4] relies on point-to-point correspondences to estimate the relative transformation of scans by minimizing the Euclidean distance error metric. The original ICP algorithm assumes that there exists a correspondence between each point of the source and model dataset. This assumption is often violated with partially overlapping scans. Some modifications to the ICP algorithm have included the maximum error cutoff metric to account for false correspondences, and did not require every point to be matched. One of the key problems is that the sparsely sampled corresponding points in two different scans often do not correspond to the same point in the 3D environment, but ICP pretends that they do. In addition, the iterative descent procedure performed in the ICP algorithm requires an initial pose of the sensor in the environment in order to converge to the global minimum.

*PhD. Candidate, Mechanical and Mechatronics Engineering, University of Waterloo
†Assistant Professor, Mechanical and Mechatronics Engineering, University of Waterloo
Many extensions to the original ICP form have been previously proposed that rely on finding unique features in the two scans, in order to improve the registration accuracy. Features based on color and intensity values [5], normals [6, 7], curvatures [8], integral volume descriptors [9], moment invariants [10], spherical harmonics [11], spin images [12], corners, lines and planes [13, 14], and combinations of the above [15] have all been suggested. However, all of these features are prone to measurement noise and cannot deal with varying sampling density within the point cloud. Locally planar surface structure was exploited by Segal et al. [16] for plane-to-plane correspondence search in generalized iterative closest point algorithm (GICP). The assumption of presence of planar structures restricts the applicability of the algorithm. Rusu et al. [17] used 16-dimensional point feature histograms that describe the local surface structure. Experimental results showing the robustness of these features to outliers and invariance to pose, sampling density, and measurement noise are lacking in the literature. In addition most of these features require extensive computational steps that invalidate their use in fast scan registration.

A relatively new approach to 3D point set registration algorithm is the normal distributions transform (NDT) that represents the underlying scene geometry as a Gaussian probability distribution [18, 19]. 3D-NDT partitions the space into disjoint volumetric cells called voxels and represents points within the voxels as a probability density function (PDF). One of the key benefits of this approach is that it forms piecewise smooth spatial representations, however the division of points into voxels results in discontinuities in the cost function, that makes it susceptible to local minima [20]. A multi-scale approach proposed by Magnusson et al. [21] minimized the discretization effects by calculating normal distributions from eight neighboring voxels. The weight of contributions from each voxel was determined by tri-linear interpolation that acts as a smoothing filter. Since PDFs have to be calculated eight times for each point, the algorithm proposed was computationally inefficient. Another multi-scale approach using K-means clustering has previously been proposed by Das et al. [22] to overcome the problem of computing analytical derivatives at voxel boundaries. The results provided are limited to 2D scans, and extension to 3D scans is not straightforward. In addition, the selection of cluster size $k$ and scale-levels is not provided. In order to address the limitations and extend the approach to 3D, Das et al. proposed a greedy clustering scheme preceded by ground segmentation [23]. The absolute translational and rotational error results presented for their approach were comparable to GICP in urban environments, however GICP failed in sparse forested environments. The requirement of classification of ground vs. non-ground points, and the assumption of spatial separation of distinct objects in the environment for greedy clustering is too restrictive.

In this paper, we extend the work of Das et al. to region growing clustering NDT (RGC-NDT) where K-means or greedy clustering has been replaced with region-growing, which requires no a-priori knowledge of cluster sizes, and makes no assumption of spatial separation of objects in the environment. A set of computationally efficient local features is computed at multiple scales and a difference between them is used to describe the point saliency. Since the features are already computed over multiple scales at the initial step,
we overcome the computationally expensive optimization performed at each successive scale level. Points adjacent to the selected seed points are merged into discrete regions according to the similarity criterion and the underlying scene geometry of the clustered regions is represented as a Gaussian probability distribution. Scan registration using distribution-to-distribution NDT matching is performed and a relative rigid body transformation between the partially overlapping scans is obtained. Experimental results demonstrate the improvement in the basin of convergence for NDT algorithm in an indoor LIDAR dataset.

The paper is organized as follows. Section 2 provides the problem formulation with a brief overview of the proposed method outlined in Section 3. Fast multi-scale difference of local saliency operator is presented in Section 4 and a description of the key steps of the algorithm including region growing clustering, and normal distribution transform is given in Sections 5. Quantitative and qualitative results for indoor scenes along with a detailed discussion, and comments on computational efficiency are provided in Section 6. Section 7 concludes the paper with directions for future work.

II. Problem Formulation

Given two 3D point sets, model set \( M = \{m_1, \cdots, m_{N_M}\} \) and data set \( D = \{d_1, \cdots, d_{N_D}\} \) where \( m_i, d_j \in \mathbb{R}^3 \) for \( i \in \{1, \cdots, N_M\}, j \in \{1, \cdots, N_D\} \), scan-to-scan registration algorithms obtain a 6-DOF relative alignment of the two scans that creates a single, globally consistent model of the environment, in order to maximize the similarity between scans. An estimate of the rigid-body transformation \( T^* = \{R, t\} \), with rotation \( R = \{R_x, R_y, R_z\} \) and translation \( t = \{t_x, t_y, t_z\} \) can be obtained by minimizing the cost function:

\[
T^* = \arg \min_{T \in S_T} \ C(M, T(D)) \tag{1}
\]

where \( S_T \) is the space of all possible rigid body transformations, \( C(M, T(D)) \) is the cost function between the model set \( M \) and the transformed data set \( T(D) \). A closed-form analytic solution to this optimization problem can be obtained in order to merge the pair of scans into a global coordinate system [24].

III. Proposed Method

Convergence basin of the NDT algorithm can be improved by getting rid of the volumetric divisions of the standard NDT by clustering points that are similar. We propose a region growing clustering NDT method (RGC-NDT) where no a-priori knowledge of the number of clusters is required. Based on the region membership function, neighboring candidate points are evaluated and subsequently merged into the current cluster, resulting in an increase in the cluster size. Normals distribution transform is applied to the clusters and used for distribution-to-distribution scan registration. The NDT algorithm is modified as follows:

1. 3D feature vectors are computed for the pair of scans at multiple scales (two scales is sufficient from experimental results). A difference operator followed by the non-maximum suppression is used to find the best candidate feature locations, as described in Section 4. The parameter selection of the search radii is dependent on the sampling density of the point sets. Strongest features typically lie at the object boundaries, and areas of high surface variation. Points with a high saliency response are unique, and do not lend themselves to clustering based on similarity. They are also susceptible to change and highly dependent on the relative position, occlusions, depth-discontinuities and sampling density. These points are discarded, thereby naturally subdividing the scan into segments that are easily clustered. This is a generalization of the segmentation of ground vs. non-ground points.

2. Points sharing a similar feature vector are then clustered using region-growing clustering as described in Section 5. As compared with the work of Das et al. [22,23] there is no prior knowledge of the number of clusters or the cluster sizes, and the optimization step is only performed once. In addition, there is no need for the pre-processing step of ground segmentation and spatial separation of distinct objects in the environment is not mandated. In the standard NDT, Gaussian distributions are abruptly truncated at the voxel boundaries resulting in discontinuities in the cost function that make it susceptible to the local minima. Clustering is able to deal with inherent limitations of gridding the metric space, and is able to segment both planar and curved surfaces of objects in indoor and outdoor environments.
3. Similar to [23,25], instead of maximizing the likelihood of a point being generated from a fixed distribution (point-to-distribution matching), we modify the NDT cost function for minimizing the $L_2$ norm between two candidate distributions (distribution-to-distribution matching), as described in Section 6. The cost function is computed for each pair of Gaussian distributions from model and data point sets, to remove all discontinuities at voxel boundaries. The reduction in the number of Gaussian components due to clustering and a compact 3D-NDT representation of the distribution results in faster run times (scoring function is faster on a few features vs all points, but points must be reliable).

IV. Multi-Scale Difference of Locally Salient Features

Points within the point cloud represent the 3D Cartesian coordinates $\{x, y, z\}$ with respect to the sensor viewpoint. Location information of these points in metric space is often insufficient to fully describe the important characteristics of the surface. Comparisons between multiple points used in the clustering application require a better metric than position in order to distinguish between points sampled from separate surfaces. In this section we describe the process of detecting locally salient features over multiple scales that aids in clustering of similar regions. The first stage detects a set of geometrical features based on empirical analysis of the 3D surface. These features are then computed over multiple scales of neighborhood search radii in order to capture characteristics of small and large scale surface structures. Difference operator followed by non-maxima suppression identifies locally salient regions in the scan.

IV.A. Feature extraction

Feature extraction is the process of transforming points from metric space to feature space such that the feature representation captures important surface details (like ridges, valleys, corners, etc.). Various methods exist for the estimation of local saliency features, however the simplest method is based on the Principal Component Analysis (PCA) of the point of interest within a neighbourhood [26].

Given a laser scan $P = \{p_1, \ldots, p_{N_p}\}$ with $p_i \in \mathbb{R}^3$ ($P = \{M\}$ for model set, and $P = \{D\}$ for data set), a set of $k$ neighbouring points $N = \{n_1, \ldots, n_k\}$ around the point of interest $p_i$ can be located within a sphere of radius $r$.

The positive semi-definite covariance matrix $C \in \mathbb{R}^{3 \times 3}$ of $N$ can be calculated as:

$$C = \frac{1}{k} \sum_{i=1}^{k} w_i \cdot (p_i - \bar{p}) \cdot (p_i - \bar{p})^T$$

where, $\bar{p}$ is the centroid of all the points in $N$, and $w_i$ represents the weighting factor for each point (typically $w_i = 1$).

Using Singular Value Decomposition (SVD), $C$ can be decomposed into the principal components of $N$ with eigenvalues $\lambda_j \in \mathbb{R}$, $j \in \{0,1,2\}$ and eigenvectors $\vec{v}_j \in \mathbb{R}^3$ forming an orthogonal frame, such that $\lambda_2 \geq \lambda_1 \geq \lambda_0 \geq 0$. These principal components describe the underlying local surface geometry and the eigenvector $\vec{v}_0$ corresponding to the smallest eigenvalue $\lambda_0$ gives an approximation of the normal $\vec{n} = \{n_x, n_y, n_z\}$ to the plane tangent to the surface through $\bar{p}$ at a distance $d_i = (p_i - \bar{p}) \cdot \vec{n}$, with $\vec{v}_1$ and $\vec{v}_2$ spanning the tangent plane $T(p)$. The calculation of the normal direction is ambiguous from PCA and is usually fixed in the direction of the sensor viewpoint.

The level of surface variation along the three eigenvectors around the point of interest can be approximated by using the eigenvalues of the covariance matrix:

$$f_0 = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}$$

$$f_1 = \frac{\lambda_1}{\lambda_0 + \lambda_1 + \lambda_2}$$

$$f_2 = \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}$$

where, $\vec{F} = \{f_0,f_1,f_2\}$ and $\vec{F} \in \mathbb{R}^3$ represents a 3D feature vector formed by normalizing the individual eigenvalues. Figure 2 shows the calculation of these features on a laser scan from Bremen dataset [27].
can be seen, feature $f_0$ has a strong response at the intersection of surfaces and weak response for points on the tangent plane to the surface, i.e. planar surfaces, whereas feature $f_1$ has a stronger response for relatively flat regions. A strong response at the depth-discontinuities (holes, occlusions, surface edges) is registered in feature $f_2$. These computationally efficient low-level local features are sufficient to coarsely define points of interest based on the level of distinctiveness of the features relative to the adjacent points called point saliency.

![Figure 2](image-url) (a) Original unlabelled point cloud. (b) Feature $f_0$ with a strong response at the intersection of surfaces. (c) Feature $f_1$ with a strong response for relatively flat regions. (c) A strong response at depth-discontinuities is registered for feature $f_2$.

IV.B. Multi-scale difference operator

Locally salient features describe the underlying geometry of the surface for a fixed support radius $r$ that captures the scale at which these features are computed. Using a fixed global $r$ can be problematic for non-uniformly sampled point clouds as the sampling densities vary locally. To alleviate this problem, the features are computed for increasing radii in the interval $[r_{\text{min}},r_{\text{max}}]$. Finer details of the surface are captured as $r$ approaches $r_{\text{min}}$, whereas an increase in radius results in blurring of the surface structure similar to the difference of Gaussians (DoG) approach for 2D images. The response of the multi-scale operator $O$ computed on the input point set $P$ for the given support radius $r$ is the set of all feature vectors $\vec{F}$, one for each point, defined as:

$$O(P,r) = \vec{F}(P,r)$$

Determining an appropriate support radius for each point in the point cloud is a computationally expensive optimization problem [28]. In this paper we propose a fast multi-scale operator $M$ similar to [6], which compares the response of the operator over multiple radii:

$$M(P,r_1,r_2) = \frac{\vec{F}(P,r_1) - \vec{F}(P,r_2)}{2} \forall r_1, r_2 \in \mathbb{R}, r_1 < r_2$$

American Institute of Aeronautics and Astronautics
The magnitude of the operator response $|M(P, r_1, r_2)|$ can be thresholded resulting in points that lie on regions with high surface variation. Figure 3 shows a laser scan (scan000) from a different viewpoint of the Bremen dataset [27] (Figure 3a). A difference of features is computed for two different support radii ($r_1 = 0.5, r_2 = 1.0$) as shown in Figure 3b. Figure 3c shows the filtered magnitude response of the difference operator.

![Figure 3. Difference operator performed on scan000 of the Bremen dataset with support radius $r_1 = 0.50$ and $r_2 = 1.0$ [27].](image)

(a) Original point cloud. (b) Difference of features (c) Filtered response with magnitude threshold $\leq 0.16$. (best viewed in colour)

V. Scan registration using Region-growing clustering NDT

Once the multi-scale feature vectors have been computed for each point in the cloud, points with low surface variability can be clustered together based on the similarity metric. In this work, region growing clustering algorithm is applied that starts the segmentation process from a random set of seed points and grows the region by connecting the neighboring points that satisfy the region membership criterion. A Gaussian distribution representation is obtained for each cluster for subsequent distribution-to-distribution NDT scan registration.

V.A. Region Growing Clustering

Region growing segmentation algorithm merges points within a certain neighborhood $N$ that share a common similarity metric $\Gamma$ into $N_c$ clusters $C = \{c_1, c_2, \cdots, c_{N_c}\}$. The goal is to assign a unique label: cluster membership $c_j \in \{1, \cdots, N_c\}$ to each point in the point cloud. The process is initiated from a random selection of $N_s$ seed points $S = \{s_1, s_2, \cdots, s_{N_s}\}$. A set of available points $A \in P \setminus S$ in the cloud is formed exclusive of these seed point locations. The neighborhood function $N$ returns a set of all adjacent points to the current seed point within a support radius $r$. Each point in the neighborhood is evaluated with respect to the current seed point based on the given constraint. For example, difference between feature vectors can
be used as the membership criterion. If the criterion is satisfied, the point along with the seed point is
merged into a cluster and subsequently removed from the set of available points. The process is repeated
until all points in the neighborhood have been evaluated against the seed point. A fresh set of seed points is
generated in the unlabelled regions once the initial set of seed points is exhausted. The growth of the regions
is terminated when all points in the cloud have been assigned a label. Points that are not assigned to any
other cluster set are assigned to the null cluster set $c_0$. A minimum and maximum cluster size threshold is
maintained to limit the region growth [29]. The method for clustering is summarized in Algorithm 1 and
the segmentation results for the LIDAR scan from Bremen dataset are provided in Figure 4.

![Region growing clustering result](image)

(a)

Figure 4. Region growing clustering result on the scan000 of the Bremen dataset. As can be seen both planar and
curved regions are nicely segmented into disjoint sets. (best viewed in colour)

V.B. Normal Distribution Transform

The normals distributions transform (NDT) grids the points into $N_v$ voxels $v_i$ and represents the data points
within as normal distributions that describes the probability of finding a surface at each point. For each
cell, the maximum likelihood estimates of the covariance $\Sigma_{v_i}$ and mean $\mu_{v_i}$ are computed, assuming the data
points within the cell have been drawn from a Gaussian distribution $\mathcal{N}(\mu_{v_i}, \Sigma_{v_i})$.

A piecewise continuous and differentiable probability density is obtained that can be used for matching
without correspondences. Stoyanov et al. extended the work of Magnusson et al. by introducing a
distribution-to-distribution cost function where the distance between the 3D-NDT representation of the
model set $M_{NDT}(M)$ and data set $D_{NDT}(M)$ is minimized to register the laser scans. $L_2$ distance metric
between the two laser scans $C_{L_2}(M, D, T^*)$ representing the cost function for distribution-to-distribution
matching is given by:

$$C_{L_2}(M, D, T^*) \approx \sum_{i=1}^{N_{M_{NDT}}} \sum_{j=i}^{N_{D_{NDT}}} \mathcal{N}(0|\mathcal{T}(\mu_i) - \mu_j, \mathcal{T}(\Sigma_i) + \Sigma_j)$$ (8)

where, $N_{M_{NDT}}$ and $N_{D_{NDT}}$ are the number of NDT components in model and data set respectively.
The objective function of the registration algorithm along with the first and second order derivatives of the
transformation function $T$ can be analytically calculated for each pair of closest Gaussian components in the
NDT representations of the model and data sets. Transformation parameters for rotation and translation
can be obtained from the Newton optimization method with line search.
Algorithm 1 Algorithm for region-growing clustering

Initialize a set of available data points:
\[ A \leftarrow \{a_1, a_2, \ldots, a_{N_a}\} \]

Initialize a set of feature vectors for each point:
\[ F \leftarrow \{f_1, f_2, \ldots, f_{N_a}\} \]

Set the threshold \( \epsilon \) for the similarity metric \( \Gamma \)

Initialize an empty cluster list:
\[ C \leftarrow \emptyset \]

while \( \{A\} \) is not empty do

Current cluster \( \{C_c\} \leftarrow \emptyset \)

Randomly select \( N_s \) seed points:
\[ S \leftarrow \{s_1, s_2, \ldots, s_{N_s}\} \]
\[ \{A\} \leftarrow \{A\} \setminus S \]

for \( i = 0 \) to \( N_s \) do

Find nearest neighbors of current seed point:
\[ \{N\} \leftarrow \Omega(s_i) \]

for \( j = 0 \) to size(\( N \)) do

if \( \Gamma(a_j, s_i) \geq \epsilon \) then

\[ \{C_c\} \leftarrow \{C_c\} \cup a_j \]

\[ \{N\} \leftarrow \{N\} \setminus a_j \]

end if

end for

end for

\[ \{C\} \leftarrow \{C\} \cup \{C_c\} \]

end while

VI. Experimental Results and Discussion

The RGC-NDT approach is evaluated using an indoor dataset and is compared against other scan registration methods such as ICP, G-ICP, and NDT. Open-source implementations of ICP, G-ICP and NDT are employed from the Point Cloud Library (PCL) [30]. The ICP and GICP algorithms were implemented with the maximum correspondence distance of 10m and the NDT algorithm was implemented with the Newton line search maximum step length of 0.1, and voxel grid resolution of 2m. For all algorithms, the maximum iterations was set to 500 and the optimization was terminated when the norm of the gradient or the norm of the step size falls below \( 10^{-6} \).

The accuracy of the RGC-NDT method was evaluated using the Rescue Arena dataset consisting of 26 laser scans with reflectance values recorded by Riegl VZ400 scanner (upto 33 million points in each scan) [27]. RIEGL tiepoints were used to manually register the scans to obtain ground truth pose estimates. Scans are filtered with a voxel grid of 0.05m to achieve a reduction in the number of points. Pair-wise scan registration was performed using every 5th scan with no initial conditions. The absolute error in rotation and translation is compared with the ground truth measurements from manual registration. Figure 5 shows the absolute translational and rotational errors. The "central box" in box-plots represents the central 50% of the data with a central line indicating the median, and the lower and upper boundary lines are at the 25% and 75% quartile of the data. Two vertical lines extending from the central box to 1.5 times the height of the central box indicate the remaining data not considered as outliers, and the red points on top represent any outliers if present.

The error distributions from Figure 5 demonstrate that RGC-NDT algorithm produces accurate results when compared to the ground truth. The large translational and rotational error in NDT is due to the convergence of the algorithm to local minima. Both ICP and G-ICP algorithms show low translational and rotational errors however, as compared with the RGC-NDT algorithm, they do not appear to be as accurate. Figure 6 displays the close-up views of the first registered scan-pair using ICP, G-ICP, NDT and RGC-NDT algorithms. Visually, the resulting aggregated registered scans using G-ICP and RGC-NDT algorithms demonstrate better alignment as compared with other algorithms.
Figure 5. Box plots for errors in transformation parameters for Rescue Arena dataset. (a) Absolute translational error. (b) Absolute rotational error.

VII. Conclusions and Future Work

This work presents a region growing clustering NDT (RGC-NDT) method for improving the convergence properties of standard NDT algorithm. Volumetric divisions and discontinuities at the voxel boundaries are removed during the optimization step. A set of computationally efficient multi-scale difference of locally salient feature vectors is computed for each point in the point cloud using principal components analysis (PCA). Adjacent points around the seed point in the local neighborhood are aggregated into clusters according to the region membership criterion based on the similarity of these feature vectors. Normal distributions transform is computed for each cluster and points within the region are represented as a probability density function (PDF) for distribution-to-distribution matching. Scan registration results from indoor laser scans validate the improvement of the basin of convergence for NDT algorithm.

References


Figure 6. Laser scan pairs scan000 (red) and scan001 (green) (a) No alignment (b) Ground Truth (c) ICP (d) GICP (e) NDT (f) RGC-NDT. (best viewed in colour)


