

Dual Transformation and Manifold Distances Voting for Outlier Rejection in Point Cloud Registration

Amit Efraim
Ben Gurion University
Beer Sheva, Israel
efraimam@post.bgu.ac.il

Joseph M. Francos
Ben Gurion University
Beer Sheva, Israel
francos@ee.bgu.ac.il

Abstract

We present a novel outlier rejection scheme for point cloud registration using $SE(3)$ voting on local transformation estimates with a dual consensus constraint. Point cloud registration is commonly performed by matching key-points in both point clouds and estimating the transformation parameters from these matches. In the presented method, each putative matching pair of points is equipped with a local transformation estimate using the Rigid Transformation Universal Manifold Embedding. Putative matching pairs with similar local estimates are then clustered together and the global transformation between point clouds is estimated for each cluster. Finally, the cluster with the majority of the votes such that the average of local transformations agrees with its associated global transformation is selected for completing the registration. This approach successfully deals with up to 99.5% outliers where state of the art fails.

1. Introduction

Registration of point cloud measurements of 3D objects has been an active research subject with a vast range of applications in computer vision, robotics, autonomous navigation and more. A point cloud \mathcal{P} is a finite set of points in \mathbb{R}^3 . In many applications these points are samples from a physical object, $\mathcal{O} \subseteq \mathbb{R}^3$ (we may think of it as a surface in 3D). Viewing point clouds as sets of samples, the registration problem may be formulated as follows: Let $\mathcal{O} \subseteq \mathbb{R}^3$ be a physical object and $T(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}$ a rigid map ($\mathbf{R} \in SO(3)$ is a rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ is a translation vector). We consider the transformed object $T(\mathcal{O}) := \{T(\mathbf{x}) : \mathbf{x} \in \mathcal{O}\}$. Let \mathcal{P} and \mathcal{Q} be two point clouds sampled from the object \mathcal{O} and the transformed

This research is supported by NSF-BSF Computing and Communication Foundations (CCF) grants, CCF-2016667, CCF-1712788 and BSF-2016667.

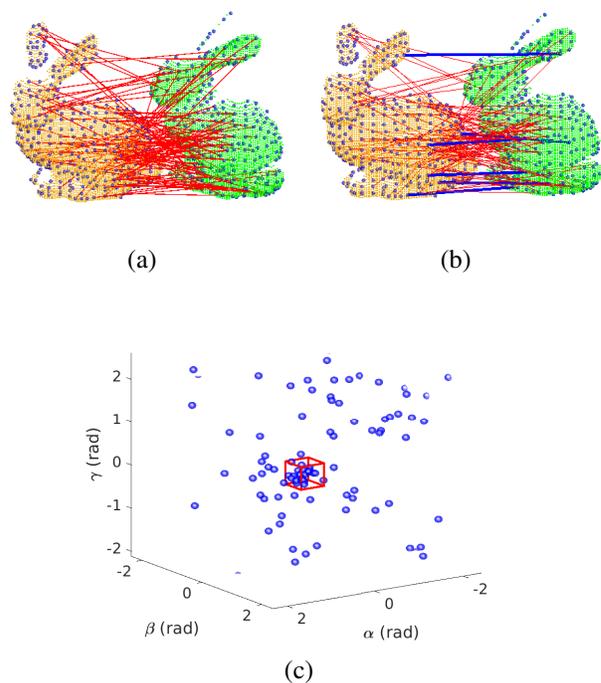


Figure 1. Outlier rejection using $SE(3)$ voting on local transformation estimates between putative matches (see, Section 5). (a) Bunny models with 45° angle between scans. Blue points mark the detected key-points, red lines mark all putative matches after initial matching. Each putative match is equipped with a local transformation estimate between matching points. (b) Bunny scans after outlier rejection using $SE(3)$ voting. Blue lines mark the detected inlier matches. (c) Scatter plot of Euler angles of the local transformation estimates. Rotation estimates of correctly matched key-points are clustered together (in red box), while false matches are randomly distributed.

object $T(\mathcal{O})$, respectively. In the registration problem, addressed in this paper, the goal is to estimate the transformation parameters \mathbf{R} and \mathbf{t} given only \mathcal{P} and \mathcal{Q} .

When there is no assumption on the initial pose of the

models undergoing registration, the standard registration pipeline is composed of three principal steps: key point detection and matching (e.g. [13, 19]); outlier rejection and initial registration (e.g. [5, 2, 16]); and finally, local optimization (e.g. [1, 14]). When sampling of the point clouds is sparse, non uniform, and noisy, which is the case in practice, key-point detection and matching is a difficult task that often results in more than 95% outliers. Due to this high outlier rate, standard methods such as RANSAC [5] might fail to find the correct consensus set (as seen in the presented experiments). State of the art methods such as TEASER++ [16] can efficiently handle very high outlier rates, however they still rely only on the geometry constraint between key-points, giving way to finding false consensus sets (see, Figure 5).

In this paper, we show that by imposing a dual consensus requirement on the putative key-point matches, outliers are accurately rejected in the presence of up to 99.5% false matches. More specifically, by adapting the Rigid Transformation Universal Manifold Embedding (RTUME) [3] to a local key-point descriptor, key-point to key-point local transformation is simultaneously estimated with the distance between the key-point descriptors. For correctly matched key-points these local estimates are also estimates of the unknown underlying transformation between the point clouds. Therefore, the local transformation estimates corresponding to correctly matched key-points form a cluster of rigid transformations among randomly distributed rigid transformations of false matches (see, Figure 1). In the presence of extremely high outlier rates however, local transformation estimates of false matches might form false clusters. Therefore, a dual consensus is imposed, *i.e.*, a set of matches must agree both on local transformations and the global transformation estimate between the matched key-points themselves. These local and global transformation estimates are independent (for false matches) and therefore the probability of finding a false set of matches that agrees on both is low, leading to a robust and accurate outlier rejection scheme.

The contribution of this work is therefore twofold:

- We introduce a new local descriptor for matching key-points on point clouds. This is the RTUME local descriptor which jointly estimates the similarity *and* the transformation between local neighborhoods of the key-points.
- Using the two information types provided by the local RTUME descriptor we propose a novel outlier rejection scheme, based on a voting scheme on the transformation manifold and a dual consensus constraint.

The rest of the paper is structured as follows. The problem is formally defined in Section 2. A short survey of related work is provided in Section 3. The technical back-

ground required for the derivation of the proposed method is presented in Section 4, where the RTUME [3] is derived and adapted to the framework of local point cloud descriptors. The proposed method for outlier rejection is then presented in Section 5. Experimental results and comparisons with existing methods are presented in Section 6. Finally, a discussion and our conclusions are provided in Section 7.

2. Problem Definition

Consider an object \mathcal{O} , and an *orbit* of equivalent objects formed by the action of the transformation group $G = \text{SE}(3)$. An observation on object \mathcal{O} will be denoted $\mathcal{P} \subset \mathbb{R}^3$ (*i.e.* the point cloud of samples from the surface of \mathcal{O}) and the set $\psi = \{\alpha \circ \mathcal{O}, \alpha \in G\}$ will denote the orbit of possible appearances of the object due to the action of the group G . There exists one such orbit for each object \mathcal{O} .

Since in practice the sampling of point clouds from different angles results with different sampled points and partially overlapping point clouds, we define subsets $\{\mathcal{P}_i\}_{i=1}^L \subset \mathcal{P}$ with corresponding orbits $\{\psi_{\mathcal{P}}^i\}_{i=1}^L$, formed by the intersection of \mathcal{P} with balls around key-points of \mathcal{P} , where the key-points are denoted by $\{\mathbf{p}_i\}_{i=1}^L$.

Given two observations of an object \mathcal{P}, \mathcal{Q} related by an unknown rigid transformation $T \in G$ *s.t.* $\mathcal{Q} = T(\mathcal{P})$, the subsets of observations $\{\mathcal{P}_i\}_{i=1}^L, \{\mathcal{Q}_k\}_{k=1}^M$ and their corresponding key-points $\{\mathbf{p}_i\}_{i=1}^L, \{\mathbf{q}_k\}_{k=1}^M$ we solve the problem of matching corresponding subsets (and thus matching corresponding key-points), such that key-points from \mathcal{P} and \mathcal{Q} and their corresponding subsets are related by the underlying transformation between observations. *i.e.* we find the *dual* consensus set, denoted by $C_{dual} \subset \mathbb{N} \times \mathbb{N}$, such that, ideally, in the absence of sampling noise,

$$\mathbf{q}_k = T(\mathbf{p}_i), \forall (i, k) \in C_{dual} \quad (1)$$

$$\mathcal{Q}_k = T(\mathcal{P}_i), \forall (i, k) \in C_{dual} \quad (2)$$

We call (1) the global consensus as it requires all the key-points to globally agree on the same transformation, and (2) the local consensus as it requires each subset to individually agree on the same transformation. Thus, the registration framework we address in this paper employs a dual-consensus constraint where we require the solution to agree both on the local and global constraints. This is in contrast to existing key-point matching algorithms where only the global consensus is employed.

Applying an adaptation of the RTUME [3] to $\{\mathcal{P}_i\}_{i=1}^L$ and $\{\mathcal{Q}_k\}_{k=1}^M$, joint estimation of the rigid transformation relating pairs of subsets of \mathcal{P} and \mathcal{Q} , and the measure of similarity between them is derived. The transformation-distance pairs of each putative match enable the derivation of the proposed dual-consensus outlier rejection method, aimed to achieve point cloud registration in the presence of very high rates of false matches. Figure 3 depicts a simple example of the complete algorithm.

3. Related Work

There is a variety of works aimed at outlier rejection and model estimation in general (e.g. [5, 15]) and for point cloud registration in particular (e.g. [4, 16]). In the following a brief summary of the most commonly practiced approaches and those related to our work is presented.

3.1. Random Sample Consensus (RANSAC)

RANSAC [5] is one of the most commonly employed outlier rejection algorithms. RANSAC is an iterative algorithm where at each iteration the minimal set of samples required for estimating the model is randomly selected and a candidate model is estimated. When a minimally selected set is composed of inliers, a larger set of inliers is expected to fit into the candidate model while outliers are expected to deviate from it. The best set in consensus with the same model is considered the inliers set and is used to estimate the model.

One of the most notable drawbacks of RANSAC is that it relies on random selection of the minimal sample set. Therefore the probability of finding a minimal set of inliers drops significantly when outlier rate is high.

3.2. Mean Shift Clustering on Rotations and Translations

In [4], registration of point clouds is performed by clustering local rotation estimates between putative matches using a mean shift algorithm. The work in [4] is closely related to the work presented in this paper as outliers are in fact rejected using the local transformation estimates. However, the work in this paper is different in two key aspects. First, the local rotation estimates in [4] are obtained from the relative rotation between local reference frames. Local reference frames have been shown to be extremely unreliable in the presence of noise, [17]. In the proposed method we employ the RTUME transformation estimate which is significantly more robust and provides the complete rigid transformation estimate rather than just the rotation. Indeed, in [4] clustering is performed to estimate the translation as well, but only after the rotation has been estimated, while using the estimated rotation. More importantly, the proposed method utilizes the dual consensus constraint (see, Section 5), effectively using both the global and local geometric information to perform the registration while rejecting outliers when global and local estimates do not agree, rather than employing only local estimates as in [4].

3.3. TEASER

In [16], registration in the presence of outliers is performed using truncated least squares estimation and semi definite relaxation (TEASER). The registration is broken into three consecutive steps by transforming the key-point

matches into translation invariant measurements and then into translation and rotation invariant measurements. First, scale is estimated and outliers are rejected using an adaptive voting scheme to solve a truncated least squares problem on the translation and rotation invariant measurements. Then rotation is estimated by solving a truncated least squares problem on the translation invariant measurements using binary programming with semi definite relaxation. Finally, the translation is estimated using a truncated least squares method and the same adaptive voting scheme used for the scale estimation. This paper and TEASER share a similar notion in the sense that local estimates are used in a voting scheme to reject outliers. However, TEASER employs only scale estimates for outlier rejection, where in the proposed method, the use of the local RTUME descriptor allows to perform voting on the entire set of parameters defining the $SE(3)$ manifold rather than on scale estimates alone.

4. The Rigid Transformation Universal Manifold Embedding and Estimation

The RTUME [3] maps functions defined on \mathbb{R}^3 to matrices, such that the transformation relating the corresponding RTUME matrices is the same as the underlying rigid transformation of coordinates between the functions (see, (4) and (7)). By mapping point clouds related by rigid transformations to corresponding RTUME matrices, the rigid transformation between point clouds and a measure of the similarity between point clouds are estimated from their RTUME matrices, as described next.

First, we present the RTUME mapping for general functions on \mathbb{R}^3 . Then we describe the required adaptations for applying the RTUME to point clouds as a local descriptor in Section 4.3.

4.1. The RTUME mapping

Let $h : \mathbb{R}^3 \mapsto \mathbb{R}$ be a finite support function and ψ_h the orbit formed by the action of $SE(3)$ on the domain of h . The RTUME matrix representation of $h(\mathbf{x})$ is given by

$$\mathbf{T}(h) = \begin{bmatrix} \int_{\mathbb{R}^3} \tilde{\mathbf{x}} w_1 \circ h(\mathbf{x}) d\mathbf{x} \\ \vdots \\ \int_{\mathbb{R}^3} \tilde{\mathbf{x}} w_N \circ h(\mathbf{x}) d\mathbf{x} \end{bmatrix} \quad (3)$$

where $\tilde{\mathbf{x}} = [1 \ \mathbf{x}^T]$ and $w_i, i = 1, \dots, N$ are measurable functions aimed at generating many compandings of the observation. (See Section 4.3.1 for the specific choice of $w_i, i = 1, \dots, N$ adopted in the numerical examples section). Note that unlike the classical moment invariant methods that use high-order moments and their nonlinear invariant functions, the RTUME representation uses low-order moments of many compandings of the observation.

Since $\{w_i\}_{i=1}^N$ are left-hand compositions, the operation of $\{w_i\}_{i=1}^N$ on h is invariant to the underlying transformation while the first order moment conserves geometric information of h , making the action of the operator \mathbf{T} *covariant* with transformation of coordinates, as described next.

Let $h(\mathbf{x}), g(\mathbf{x}) \in \psi_h$ be two functions from the same orbit such that

$$h(\mathbf{x}) = g(\mathbf{R}\mathbf{x} + \mathbf{t}) \quad (4)$$

By a change of variables $\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{t}$ it follows that $\tilde{\mathbf{y}} = \tilde{\mathbf{x}}\mathbf{D}(\mathbf{R}, \mathbf{t})$ where $\tilde{\mathbf{y}} = [1 \ \mathbf{y}^T]$ and

$$\mathbf{D}(\mathbf{R}, \mathbf{t}) = \begin{bmatrix} 1 & \mathbf{t}^T \\ \mathbf{0} & \mathbf{R}^T \end{bmatrix} \quad (5)$$

is the matrix representation of $SE(3)$ in homogeneous coordinates using matrix multiplication on the right. The RTUME mapping of $g(\mathbf{y})$ is therefore

$$\begin{aligned} \mathbf{T}(g) &= \\ \begin{bmatrix} \int_{\mathbb{R}^3} \tilde{\mathbf{y}} w_1 \circ g(\mathbf{y}) d\mathbf{y} \\ \vdots \\ \int_{\mathbb{R}^3} \tilde{\mathbf{y}} w_N \circ g(\mathbf{y}) d\mathbf{y} \end{bmatrix} &= \begin{bmatrix} \int_{\mathbb{R}^3} \tilde{\mathbf{x}} \mathbf{D}(\mathbf{R}, \mathbf{t}) w_1 \circ g(\mathbf{R}\mathbf{x} + \mathbf{t}) d\mathbf{x} \\ \vdots \\ \int_{\mathbb{R}^3} \tilde{\mathbf{x}} \mathbf{D}(\mathbf{R}, \mathbf{t}) w_N \circ g(\mathbf{R}\mathbf{x} + \mathbf{t}) d\mathbf{x} \end{bmatrix} \\ &= \begin{bmatrix} \int_{\mathbb{R}^3} \tilde{\mathbf{x}} w_1 \circ h(\mathbf{x}) d\mathbf{x} \\ \vdots \\ \int_{\mathbb{R}^3} \tilde{\mathbf{x}} w_N \circ h(\mathbf{x}) d\mathbf{x} \end{bmatrix} \mathbf{D}(\mathbf{R}, \mathbf{t}) = \mathbf{T}(h) \mathbf{D}(\mathbf{R}, \mathbf{t}) \quad (6) \end{aligned}$$

The RTUME matrices $\mathbf{T}(h)$ and $\mathbf{T}(g)$ constructed from $h(\mathbf{x})$ and $g(\mathbf{x})$ as in (3) are therefore related by the relation

$$\mathbf{T}(h) = \mathbf{T}(g) \mathbf{D}^{-1}(\mathbf{R}, \mathbf{t}) \quad (7)$$

Since $\mathbf{T}(h)$ and $\mathbf{T}(g)$ are related by an invertible transformation that is a re-expression of the rigid transformation relating the functions h and g , we say that the basis $\mathbf{T}(g) \mathbf{D}^{-1}(\mathbf{R}, \mathbf{t})$ is *covariant* with the rigid transformation. Hence it provides a method for estimating the transformation that relates any two functions related by a rigid transformation of coordinates.

4.2. Joint Estimation of the Distance and Transformation between RTUME matrices

Using (7) we have that for any two functions taken from the same orbit, their RTUME representations are linearly related. Hence, for a given choice of $\{w_i\}_{i=1}^N$ their corresponding RTUME matrices form an equivalence class in the space of $N \times 4$ matrices. The distance between RTUME equivalence classes is therefore a representation of the distance between orbits of functions. Let $h(\mathbf{x})$ and $g(\mathbf{x})$ be two functions (not necessarily from the same orbit) and denote their corresponding RTUME matrices by $\mathbf{H} = \mathbf{T}(h)$

and $\mathbf{G} = \mathbf{T}(g)$. Let $[\mathbf{H}]$ and $[\mathbf{G}]$ denote the equivalence classes of RTUME matrices constructed from h and g . The RTUME distance between the corresponding RTUME matrices is given by

$$d(\mathbf{H}, \mathbf{G}) = \min_{\substack{\mathbf{Q} \in [\mathbf{H}] \\ \mathbf{P} \in [\mathbf{G}]}} \|\mathbf{Q} - \mathbf{P}\|_F^2 \quad (8)$$

Alternatively, using (7), the elements of the equivalence classes $[\mathbf{H}]$ and $[\mathbf{G}]$ can be written explicitly using the representatives \mathbf{H} and \mathbf{G} as follows

$$d(\mathbf{H}, \mathbf{G}) = \min_{\mathbf{D}_H, \mathbf{D}_G \in SE(3)} \|\mathbf{H} \mathbf{D}_H - \mathbf{G} \mathbf{D}_G\|_F^2 \quad (9)$$

A closed form solution for (9) is found as described next. First, the operation of \mathbf{D}_H and \mathbf{D}_G is represented as a composition of rotation and translation operations. Let \mathbf{h}_{m0} denote the leftmost column of \mathbf{H} , *i.e.*, the zero order moment in (3) and \mathbf{H}_{m1} the three rightmost columns of \mathbf{H} , *i.e.*, the first order moments in (3). Similarly define \mathbf{g}_{m0} and \mathbf{G}_{m1} with respect to \mathbf{G} . Also, let $\mathbf{R}_H, \mathbf{R}_G \in SO(3)$ and $\mathbf{t}_H, \mathbf{t}_G \in \mathbb{R}^3$ be the rotation and translation components of \mathbf{D}_H and \mathbf{D}_G . An equivalent formulation of (9) is given by

$$\begin{aligned} \|\mathbf{H} \mathbf{D}_H - \mathbf{G} \mathbf{D}_G\|_F^2 &= \|\mathbf{h}_{m0} - \mathbf{g}_{m0}\|_F^2 + \\ &\|\mathbf{H}_{m1} \mathbf{R}_H + \mathbf{h}_{m0} \mathbf{t}_H - \mathbf{G}_{m1} \mathbf{R}_G - \mathbf{g}_{m0} \mathbf{t}_G\|_F^2 \quad (10) \end{aligned}$$

The first term in (10) is independent of the transformation and therefore doesn't affect the minimization problem. By expanding the second term in (10) using the definition of the Frobenius norm we have the following

$$\begin{aligned} \|\mathbf{H}_{m1} \mathbf{R}_H + \mathbf{h}_{m0} \mathbf{t}_H - \mathbf{G}_{m1} \mathbf{R}_G - \mathbf{g}_{m0} \mathbf{t}_G\|_F^2 &= \\ \|\mathbf{H}_{m1}\|_F^2 + \|\mathbf{G}_{m1}\|_F^2 & \\ - 2 \operatorname{tr}(\mathbf{G}_{m1} \mathbf{R}_G \mathbf{R}_H^T \mathbf{H}_{m1}^T) & \\ - 2 \operatorname{tr}(\mathbf{h}_{m0} \mathbf{t}_H \mathbf{R}_G^T \mathbf{G}_{m1}^T) + 2 \operatorname{tr}(\mathbf{g}_{m0} \mathbf{t}_G \mathbf{R}_G^T \mathbf{G}_{m1}^T) & \\ - 2 \operatorname{tr}(\mathbf{g}_{m0} \mathbf{t}_G \mathbf{R}_H^T \mathbf{H}_{m1}^T) + 2 \operatorname{tr}(\mathbf{h}_{m0} \mathbf{t}_H \mathbf{R}_H^T \mathbf{H}_{m1}^T) & \\ + \operatorname{tr}(\mathbf{g}_{m0} \mathbf{t}_G \mathbf{t}_G^T \mathbf{g}_{m0}^T) + \operatorname{tr}(\mathbf{h}_{m0} \mathbf{t}_H \mathbf{t}_H^T \mathbf{h}_{m0}^T) & \\ - 2 \operatorname{tr}(\mathbf{g}_{m0} \mathbf{t}_G \mathbf{t}_H^T \mathbf{h}_{m0}^T) & \quad (11) \end{aligned}$$

The minimization of (11) is performed by first finding the translations \mathbf{t}_H and \mathbf{t}_G that minimize the above expression as functions of \mathbf{R}_G and \mathbf{R}_H . The translations are found by evaluating the derivatives of (11) with respect to \mathbf{t}_H and \mathbf{t}_G . Then, substituting back these translations into (11) we are left with an expression of the general form $\operatorname{tr}(\mathbf{A}^T \mathbf{R})$ that has to be maximized in order to minimize (11). This is a well studied problem and is solved following [12].

For RTUME matrices evaluated from functions taken from the same orbit, $\hat{\mathbf{D}}_H, \hat{\mathbf{D}}_G \in SE(3)$ that minimize (9) form an estimate for the relative transformation between \mathbf{H}

and \mathbf{G} . It follows from (7) that this is also an estimate for the underlying transformation of coordinates

$$\widehat{\mathbf{D}}_{\mathbf{H},\mathbf{G}} = \widehat{\mathbf{D}}_{\mathbf{H}}^{-1} \widehat{\mathbf{D}}_{\mathbf{G}} \quad (12)$$

This means that the distance between RTUME matrices is jointly estimated with the underlying transformation of coordinates. When used for local key-point registration, this joint estimation enables the voting scheme for outliers rejection described in Section 5.

4.3. The RTUME as local key-point descriptor

Consider the problem defined in Section 2. In the following, we describe how the RTUME is adapted for point clouds in general and as a local key-point descriptor in particular.

4.3.1 The RTUME for point clouds

Since point clouds are sets of coordinates in 3-D with no functional relation imposed on them a necessary step in adapting the RTUME framework for point cloud processing is to define a function that assigns each point in the cloud with a value, invariant to the action of the transformation group. Existing point cloud descriptors can be utilized to this end (e.g. [13, 7, 8]) by assigning every point of the point cloud with the descriptor value evaluated around it. (See, Figure 2).

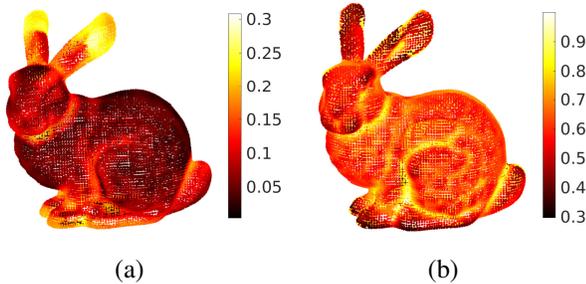


Figure 2. Bunny model from the Stanford 3D scanning repository, imposed with a function, invariant to rigid transformations. (a) Using the surface variation descriptor from [8]. (b) Using the volume curvature descriptor [7].

The second challenge in adapting the RTUME for point clouds is in the integration. Even with the function defined on each point of the point cloud, the Lebesgue measure of the point cloud in \mathbb{R}^3 is zero. Therefore, integration is performed using the counting measure instead, resulting with sums. Let $\mathcal{P} \subset \mathbb{R}^3$ be a point cloud, and $h(\mathbf{x}), \mathbf{x} \in \mathcal{P}$ the function defined on the point cloud. Then, the RTUME

mapping for point clouds is given by

$$\mathbf{T}(h) = \begin{bmatrix} \frac{1}{|\mathcal{P}|} \sum_{\mathbf{x} \in \mathcal{P}} \tilde{\mathbf{x}} w_1 \circ h(\mathbf{x}) \\ \vdots \\ \frac{1}{|\mathcal{P}|} \sum_{\mathbf{x} \in \mathcal{P}} \tilde{\mathbf{x}} w_N \circ h(\mathbf{x}) \end{bmatrix} \quad (13)$$

The factor of $\frac{1}{|\mathcal{P}|}$ was added to compensate for the possibly different sampling densities, which translates to a scale factor between the RTUME matrices in (7) unless normalizing by $|\mathcal{P}|$.

Following [18] the set $\{w_i\}_{i=1}^N$ is chosen as the set of indicator functions, uniformly quantizing the values of the function h defined on the point cloud. This choice yields level-set functions, computed at each quantization level in an observation. These level-set functions serve as a basis for the invariant subspaces in RTUME.

4.3.2 The local RTUME point cloud descriptor

Using the notation defined in Section 2, let $h(\mathbf{x})$ denote a function defined on some point cloud \mathcal{P} . Given a key point $\mathbf{p}_k \in \mathcal{P}$ and its neighborhood \mathcal{P}_k , the *local RTUME descriptor* of \mathbf{p}_k is denoted by $\mathbf{H}_{\mathcal{P}_k}$. It is evaluated using a local adaptation of (13), such that the sums are evaluated locally on the subset \mathcal{P}_k .

Given \mathcal{P} and \mathcal{Q} , with key-points $\{\mathbf{p}_i\}_{i=1}^L$ and $\{\mathbf{q}_k\}_{k=1}^M$, and the local subsets $\{\mathcal{P}_i\}_{i=1}^L$ and $\{\mathcal{Q}_k\}_{k=1}^M$, corresponding sets $\{\mathbf{H}_{\mathcal{P}_i}\}_{i=1}^L$ and $\{\mathbf{H}_{\mathcal{Q}_k}\}_{k=1}^M$ of local RTUME descriptors are evaluated.

For these sets of key-points and their neighborhoods, denote the estimate of the local transformation between $\mathbf{H}_{\mathcal{P}_i}$ and $\mathbf{H}_{\mathcal{Q}_k}$ evaluated using (12) by $\widehat{\mathbf{D}}_{\mathcal{P}_i, \mathcal{Q}_k}$. The distances $d(\mathbf{H}_{\mathcal{P}_i}, \mathbf{H}_{\mathcal{Q}_k})$ defined in (9) are invariant to rigid transformations of the point clouds, and are therefore suitable for estimating the local geometric similarity between key-point neighborhoods for an initial matching (e.g. by nearest neighbor search with respect to descriptor distances), while the local transformation estimates are used next in the proposed outlier rejection scheme.

5. Outlier Rejection Using Manifold Voting and Dual Consensus

Let us, for simplicity of notation, denote the putative matches and their corresponding local transformation estimates by $\{\mathbf{p}_i, \mathbf{q}_i\}_{i=1}^N$ and $\{\widehat{\mathbf{D}}_i\}_{i=1}^N$, where N is the number of putative matches. Also, denote the rotation quaternion of the i -th local transformation estimate by $\hat{\mathbf{q}}_i$ and the translation element of the i -th local transformation estimate by $\hat{\mathbf{t}}_i$. For each local estimate, the voting is performed by evaluating its distances to all the other local estimates with respect to the rotation distance and the translation distance. Next we

obtain those local estimates that are within a certain threshold as the consensus set of that local estimate. The rotation distance can be any metric on $SO(3)$ (see [11] for a survey). In our experiments we use the following metric on rotation quaternions

$$\Phi(q_1, q_2) = \arccos(|q_1 \cdot q_2|) \quad (14)$$

More formally, the consensus set for the i -th local estimation, by $SE(3)$ voting, is defined as

$$C_i = \{k \in \mathbb{N} \mid \Phi(\hat{\mathbf{q}}_i, \hat{\mathbf{q}}_k) < \alpha \cap \|\hat{\mathbf{t}}_i - \hat{\mathbf{t}}_k\|_2 < \beta\} \quad (15)$$

where α and β are the thresholds for rotation and translation. The set of estimated inliers by voting is the largest set of indices agreeing on the transformation up to the defined threshold, *i.e.*

$$C_{\text{voting}} = \underset{C \in \{C_i\}_{i=1}^N}{\text{argmax}} |C| \quad (16)$$

While this voting scheme provides an acceptable solution for moderate to high outlier rates, in the case where outlier rates are extremely high the following dual consensus is employed: For each consensus set, the global transformation between the key-points associated with that consensus set is estimated using the least squares estimator, [10]. Let $\hat{\mathbf{D}}_{C,i}$ be the estimated global transformation from the matches in the i -th consensus set. Let $\bar{\mathbf{D}}_{C,i}$ be the average of local transformations in the i -th consensus set. The average rotation of local transformations is obtained using the L_2 -mean on $SO(3)$ from [9] while the translation is obtained by an average on \mathbb{R}^3 . A voting consensus set is accepted only if the rotation and translation distances between $\bar{\mathbf{D}}_{C,i}$ and $\hat{\mathbf{D}}_{C,i}$ are below a certain threshold. *i.e.* the *dual* consensus set is

$$C_{\text{dual}} = \underset{C \in \{C_i\}_{i=1}^N}{\text{argmax}} |C| \quad \text{s.t. } \Phi(\hat{\mathbf{q}}_C, \bar{\mathbf{q}}_C) < \alpha \text{ and } \|\hat{\mathbf{t}}_C - \bar{\mathbf{t}}_C\|_2 < \beta \quad (17)$$

Where $\hat{\mathbf{q}}_C$ and $\hat{\mathbf{t}}_C$ are the rotation quaternion and translation corresponding to $\hat{\mathbf{D}}_C$, and $\bar{\mathbf{q}}_C$ and $\bar{\mathbf{t}}_C$ are the rotation quaternion and translation corresponding to $\bar{\mathbf{D}}_C$, while α and β are thresholds for determining the similarity required between the local estimates average and the global estimate of the transformation.

The threshold values depend on the nature of the problem being solved. Our experimental results indicate that when the observed 3-D scans are accurate, the threshold for rotation clustering and dual consensus can be chosen to be as small as 3° . With more challenging data such as point clouds obtained from handheld RGB-D cameras or surface from motion, a more conservative threshold of $15^\circ - 20^\circ$ is employed, as the majority of outliers is still rejected even

with larger thresholds. Setting the threshold value on the translation is more complex as the translation error is coupled with the rotation error. Figure 3 depicts a simple example of the complete registration pipeline.

6. Experimental Results

In the following we present some experimental evaluation of the proposed method both in terms of robustness to outliers and in terms of the resulting registration accuracy. For registration accuracy we compare the results with a Matlab implementation of RANSAC [15] using a maximum of 10,000 trials and TEASER++ [16]. We use the Stanford Bunny model from the Stanford 3D scanning repository to create partially overlapping point clouds, transformed by random rigid transformations. Then, the surface variation descriptor [8] is used to define a function on each point cloud (see, Figure 2). Key points are uniformly sampled on each point cloud. The local RTUME descriptor is then constructed for each key point as described in Section 4.3.2. Initial matches are then acquired by choosing the N matches with lowest RTUME distance (9), where N is chosen per experiment to control the outlier rate. By increasing N , more false matches are allowed. Then, using the ground truth transformation the numbers of correct and false matches are selected to provide the desired outlier ratio for the test. For each outlier rate the experiment was repeated 100 times with a random transformation and matches, at each. Figure 5 depicts the experimental setup and results for a single realization of the data.

6.1. $SE(3)$ voting vs. dual consensus

In this part of the experiments we compare the performance of the $SE(3)$ voting with and without the dual consensus in terms of matching precision (defined by the ratio between the number of correct matches and the total number of matches). RANSAC is also included for comparison. From Table 1 it is concluded that when outlier rates are very high (more than 95%), randomly distributed local estimates of outliers may create false clusters of transformations as evident by the lower accuracy obtained when $SE(3)$ dual-consensus voting is not employed.

6.2. Registration Accuracy

In this part the registration accuracy is evaluated in the presence of varying outlier rates. For registration with the proposed method, we employ a two-stage procedure where RANSAC is employed following an outlier rejection stage implemented by the $SE(3)$ dual-consensus voting. Since the number of false matches is drastically reduced by the dual-consensus voting procedure, the residual outlier rate is very low as indicated in Table 1, and hence, the probability for successful registration is significantly increased. To further evaluate the registration accuracy we adopt the registra-

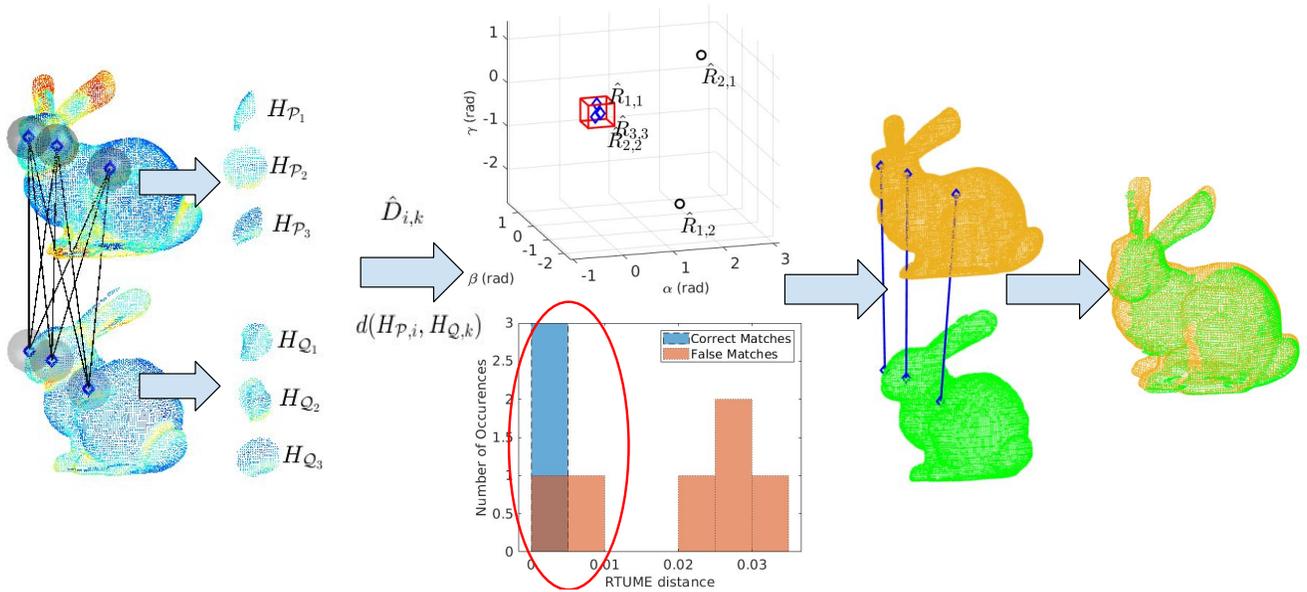


Figure 3. A simple example of the proposed registration scheme. On the left, two scans of the bunny models with three detected key-points. Note that the models are imposed with the surface variation descriptor as described in Section 4.3.1). A ball with a radius of 10 times the point cloud resolution is used to crop patches around each key-point (seen to the right of the scanned models). For each key-point and its neighborhood patch, corresponding sets $\{\mathbf{H}_{\mathcal{P}_i}\}_{i=1}^3$ and $\{\mathbf{H}_{\mathcal{Q}_k}\}_{k=1}^3$ of local RTUME descriptors are obtained as described in Section 4.3.2. Then the RTUME distances and the corresponding transformations are estimated for all possible pairs using Section 4.2. In this example key-point pairs with RTUME distance below 0.01 were selected as putative matches. As seen in the RTUME distances histogram, this threshold results with three inliers and two outliers. Then, using the dual-consensus $SE(3)$ voting, described in Section 5, the two outliers are rejected. The cluster of estimated rotations is seen contained in the red box of the rotations scatter plot, containing the three inliers ($\hat{\mathbf{R}}_{i,k}$ denotes the rotation element of $\hat{\mathbf{D}}_{i,k}$). The cluster of translations is omitted for simpler illustration. In this example the largest cluster is indeed the correct one. In terms of dual-consensus, this cluster is accepted since the errors between the average of local estimates and the global estimate are approximately 5° for the rotation and 7 times the point cloud resolution for the translation. Then, using this estimated set of matches the transformation is estimated and the two scans are registered as depicted on the right.

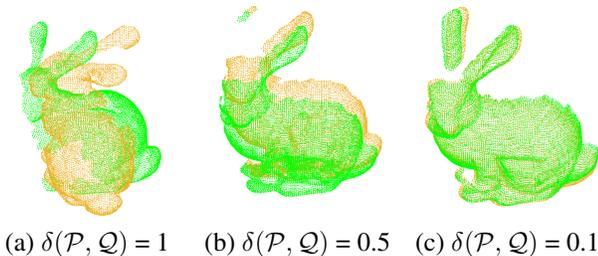


Figure 4. Examples of different point cloud poses and the associated values of the metric used for evaluating registration accuracy.

tion accuracy metric from [6]. The advantage of this metric is that registration accuracy is evaluated using a single number rather than by rotation and translation separately. Given the same point cloud in different poses \mathcal{P} and \mathcal{Q} , with M points $\{\mathbf{p}_i\}_{i=1}^M$ and $\{\mathbf{q}_i\}_{i=1}^M$ and $\bar{\mathbf{p}}$ the centroid of \mathcal{P} the distance between \mathcal{P} and \mathcal{Q} is evaluated by

$$\delta(\mathcal{P}, \mathcal{Q}) = \frac{1}{M} \sum_{i=1}^M \frac{\|\mathbf{p}_i - \mathbf{q}_i\|_2}{\|\mathbf{p}_i - \bar{\mathbf{p}}\|_2} \quad (18)$$

Outliers	90%	95%	99%	99.5%
$SE(3)$ voting	0.92	0.91	0.57	0.38
$SE(3)$ dual-consensus voting	0.93	0.91	0.86	0.49
RANSAC	0.99	0.90	0.17	0.03

Table 1. Mean precision of matches after outlier rejection for various outlier rates using $SE(3)$ voting with and without dual consensus, (see, Section 5) compared to RANSAC. $SE(3)$ voting with dual-consensus scheme performs significantly better than simple $SE(3)$ voting and RANSAC when outlier rates are higher than 95%.

Figure 4 depicts some examples of the relative pose between point clouds and the associated values of $\delta(\mathcal{P}, \mathcal{Q})$. We employ this metric to evaluate the registration accuracy between the observation point cloud transformed by the ground truth transformation and the estimated transformation. As seen in Table 2 the advantage of the proposed method is in its robustness to high outlier rates such that registration accuracy remains approximately the same despite the increase in outlier rates.

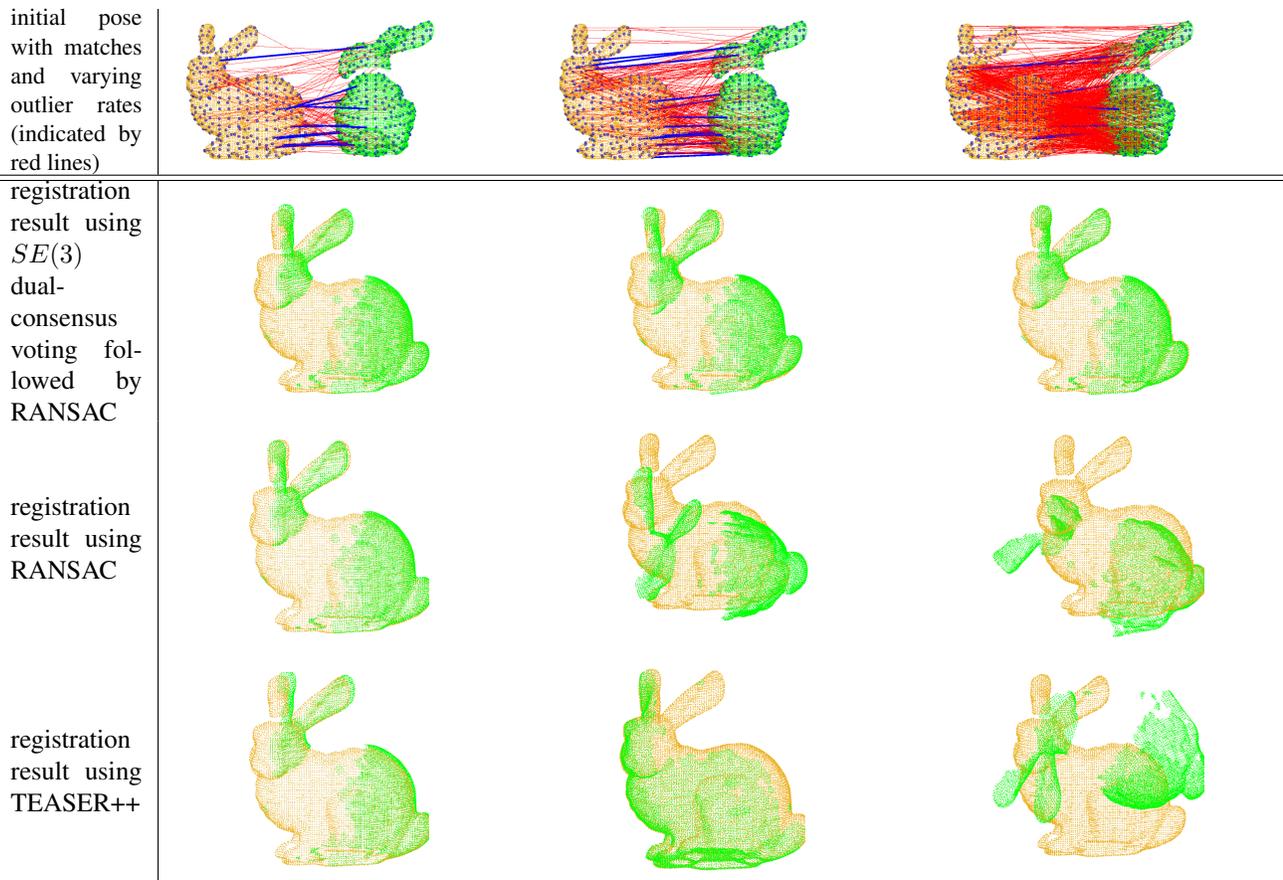


Figure 5. Experimental setup. Top row, a single realization of the source and target point clouds is presented for various outlier ratios (from left to right: 75%, 95% and 99% outliers). Red lines depict the false matches while blue lines depict the matches found using $SE(3)$ dual-consensus voting. Bottom rows: registration result using $SE(3)$ dual-consensus voting followed by RANSAC, RANSAC alone, and TEASER++. Note that with 95% and 99% outliers RANSAC and TEASER++ failed the registration while $SE(3)$ dual-consensus voting followed by RANSAC is successful.

Outliers	90%	95%	99%	99.5%
$SE(3)$ dual-consensus voting	0.06	0.08	0.05	0.23
TEASER++	0.03	0.04	0.96	1.25
RANSAC	0.05	0.14	0.65	0.47

Table 2. Mean registration accuracy evaluated using (18) with various outlier rates, comparing the proposed method, RANSAC and TEASER++. When outlier rates are very high the proposed $SE(3)$ voting with dual consensus outperforms the alternatives.

7. Discussion and Conclusions

We presented a novel outliers rejection method for point cloud registration using a dual consensus paradigm. Local transformation estimates between putative matches are computed jointly with the distance between key-point descriptors using the local RTUME descriptor. Since local transformations of correctly matched key-points are expected to be the same, the local transformations are used for

majority vote in order to reject matches that deviate from the majority of local estimates. To improve robustness when high outlier rates are present, a set of matches is accepted only if the average of local transformations agrees with the global transformation estimated from the matching points.

In the presented experiments, the proposed method proved to be robust even for an outlier rate of 99.5%. However, the experiments were conducted on synthetic data and further experimental work is required using real data (e.g. [6]) to further evaluate and validate the performance of the proposed method.

References

- [1] Paul J Besl and Neil D McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992. 2
- [2] Alvaro Parra Bustos and Tat-Jun Chin. Guaranteed outlier removal for point cloud registration with correspondences.

- IEEE Trans. Pattern Anal. Mach. Intell.*, 40(12):2868–2882, Dec 2018. 2
- [3] Amit Efraim and Joseph M Francos. The universal manifold embedding for estimating rigid transformations of point clouds. *IEEE Int. Conf. Acoust., Speech, Signal Processing, 2019*. 2, 3
- [4] Ido Haim Ferencz and Ilan Shimshoni. Registration of 3d point clouds using mean shift clustering on rotations and translations. *2017 Int. Conf. 3D Vis. (3DV)*, pages 374–382, 2017. 3
- [5] Martin A Fischler and Robert C Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. 2, 3
- [6] Simone Fontana, Daniele Cattaneo, Augusto L. Ballardini, Matteo Vaghi, and Domenico G. Sorrenti. A benchmark for point clouds registration algorithms. *Robotics and Autonomous Systems*, 140:103734, 2021. 7, 8
- [7] Natasha Gelfand, Niloy J Mitra, Leonidas J Guibas, and Helmut Pottmann. Robust global registration. In *SGP 2005: Third Eurographics Symposium on Geometry processing*, pages 197–206. Eurographics Association, 2005. 5
- [8] Timo Hackel, Jan Wegner, and Konrad Schindler. Fast semantic segmentation of 3d point clouds with strongly varying density. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3:177–184, 06 2016. 5, 6
- [9] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, Jul 2013. 6
- [10] Berthold KP Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, Apr 1987. 6
- [11] Du Huynh. Metrics for 3d rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35:155–164, 10 2009. 6
- [12] Andriy Myronenko and Xubo B. Song. On the closed-form solution of the rotation matrix arising in computer vision problems. *ArXiv*, abs/0904.1613, 2009. 4
- [13] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE International Conference on Robotics and Automation*, pages 3212–3217, May 2009. 2, 5
- [14] Thomas Schmiedel, Erik Einhorn, and Horst-Michael Gross. Iron: A fast interest point descriptor for robust ndt-map matching and its application to robot localization. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3144–3151, Sep. 2015. 2
- [15] Philip HS Torr and Andrew Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Comput. Vis. Image Und.*, 78(1):138 – 156, 2000. 3, 6
- [16] Heng Yang, Jingnan Shi, and Luca Carlone. Teaser: Fast and certifiable point cloud registration. *IEEE Transactions on Robotics*, 37(2):314–333, 2021. 2, 3, 6
- [17] Jiaqi Yang, Yang Xiao, and Zhiguo Cao. Toward the repeatability and robustness of the local reference frame for 3d shape matching: An evaluation. *IEEE Trans. Image Process.*, 27(8):3766–3781, Aug 2018. 3
- [18] Ziv Yavo, Yuval Haitman, Joseph M Francos, and Louis L Scharf. Matched manifold detection for group-invariant registration and classification of images. *IEEE Trans. Sig. Process.*, July 2021. 5
- [19] Yu Zhong. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *2009 IEEE 12th Int. Conf. Comput. Vis. Workshops, ICCV Workshops*, pages 689–696, Sept 2009. 2