THE UNIVERSAL MANIFOLD EMBEDDING FOR ESTIMATING RIGID TRANSFORMATIONS OF POINT CLOUDS

Amit Efraim and Joseph M. Francos

Electrical and Computer Engineering Department
Ben-Gurion University, Beer-Sheva, Israel

ABSTRACT

We present a closed form solution to the problem of registration and detection of dense 3-D point clouds undergoing unknown rigid deformations. The solution is obtained by adapting the general framework of the universal manifold embedding (UME) to the case where the deformations the object may undergo are rigid. The UME nonlinearly maps functions (*e.g.* , images, 3D models) related by geometric transformations of coordinates to the same linear subspace of some Euclidean space. Therefore registration, matching and classification are solved as linear problems in a lower dimensional space. In this paper we extend the UME framework to the special case where it is *a-priori* known that the geometric transformations are rigid (*e.g.* pose change of a 3-D rigid object). We further demonstrate the applicability of the methodology for the registration of 3-D point clouds. In the case where point correspondences are unknown, the majority of existing methods for registering 3-D point clouds are based on iteratively finding a transformation which minimizes some distance between the object and a model. The method proposed in this paper is notably different as registration is performed using a closed form solution that employs the UME low dimensional representation of the shapes to be registered.

***Index Terms*—** Affine Transformation, Rigid Transformation, Registration, Parameter Estimation.

1. INTRODUCTION

Registration of point cloud measurements of rigid 3D objects has been an active research subject with suggested solutions ranging from closed form solutions to iterative and numerical solutions. Applications for such registration techniques are vast, from problems in computer vision, robotics and navigation, to medical imagery and instruments. With the advancement in recent years of highly accurate dense 3D reconstruction from monocular cameras [1, 2, 3] and the increasing availability of cameras acquiring range data, the potential applications of such techniques have grown even larger. 3-D shape registration can be categorized into global and local solutions. Global solutions aim at aligning two or more data sets while taking the complete available data sets into account while local solutions are based on extracting and matching key points in order to find point correspondences.

Global approaches utilize all the available data so potentially they may achieve better results, but they also tend to suffer greatly when observed shapes are incomplete or when only a partial overlap is available. On the other hand, local solutions are robust in terms of data incompleteness and partial overlaps, but utilize less data and are prone to correspondence errors and key-point inaccuracies. The majority of existing global solutions is of iterative nature, where a distance function is defined from points in the observation either to assumed correspondence points in the model [4, 5] or to the surface of the model [4, 6, 7]. They then iteratively find a transformation which minimize this distance. The Iterative Closest Point algorithm (ICP) [4, 5] is by far the most widely used algorithm. The ICP algorithm constructs point correspondences based on spatial proximity followed by a registration step (*e.g.* [8]). By definition, the ICP and other iterative methods (*e.g.* [6, 7]) rely on a good initial alignment, otherwise registration may converge to a local minimum. The method proposed in this paper is notably different as registration is performed using a closed form solution that employs the UME low dimensional representation of the shapes to be registered.

More specifically, in this paper we show that enforcing a rigid deformation constraint on the Universal Manifold Embedding (UME) [9], derived for the case where the geometric transformation is affine, results in accurate and computationally efficient closed form solution to the problem of registration of 3-D point clouds. Moreover, the presented method is independent of the initial pose of the point clouds and data sampling rates. Since the UME low dimensional representation is also independent of the data dimensions it can handle very large data sets, which is a major limitation for existing iterative methods. When only partial overlap is available, an hierarchical search is employed in order to jointly detect the object and estimate the transformation it undergoes relative to the larger reference model, as presented in section 4. In the presented example we demonstrate the applicability of the method in the framework of 3-D surface registration where a

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

3-D point cloud obtained using multiview 3-D reconstruction is registered to a known reference model. The results are compared to the results achieved by the ICP algorithm. The rest of this paper is structured as follows. In section 2 a brief technical background required for the derivation of the proposed solution is provided. In section 3 the derivation of the UME for the special case of a rigid deformation is presented and in section 4 experimental results are provided.

2. BACKGROUND

Next we briefly review the two methodologies that form the basis of the framework presented in this paper. These are the UME [9] and the closed form solution for registration of point clouds from known correspondences, [10].

2.1. Closed form solution for registration of point clouds from known correspondences

In [8, 10] a closed form solution for the problem of recovering the transformation between two sets of points in different Cartesian coordinate systems is derived. More specifically,

$$\mathbf{r}_{r,i} = s\mathbf{R}\mathbf{r}_{l,i} + \mathbf{r}_0,\ i = 1, ..., n \tag{1}$$

where $\{\mathbf{r}_{r,i}\}_{i=1}^{n}$ is the "right hand" (or reference) set of points mapped to the "left hand" (or observed) set of points $\{\mathbf{r}_{l,i}\}_{i=1}^{n}$ by a transformation defined by s, $\mathbf{R}$ and $\mathbf{r}_0$. The goal then is to find $\mathbf{R}, \mathbf{r}_0$ and s such that we minimize

$$\sum_{i=1}^{n} ||\mathbf{e}_i||^2 = \sum_{i=1}^{n} ||\mathbf{r}_{r,i} - s\mathbf{R}\mathbf{r}_{l,i} - \mathbf{r}_0||^2 \tag{2}$$

In [10] it is shown that the translation $\mathbf{r}_0$ that minimizes the sum of squared errors is the difference between the centroid of $\{\mathbf{r}_{r,i}\}_{i=1}^{n}$ and the rotated centroid of $\{\mathbf{r}_{l,i}\}_{i=1}^{n}$, *i.e.* $\mathbf{r}_0 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}_{r,i} - \frac{1}{n}\mathbf{R}\sum_{i=1}^{n}\mathbf{r}_{r,i}$. The rotation that minimizes the sum of squared errors is shown to be the matrix maximizing $\mathrm{tr}\{\mathbf{R}^T\mathbf{M}\}$, where $\mathbf{M} = \sum_{i=1}^{n}\mathbf{r'}_{r,i}, \mathbf{r'}_{l,i}^T$ (here $\mathbf{r'}_{r,i}, \mathbf{r'}_{l,i}$ represent the centered version of $\mathbf{r}_{r,i}, \mathbf{r}_{l,i}$). $\mathbf{M}$ is shown to have a decomposition $\mathbf{M} = \mathbf{U}\mathbf{S}$ where $\mathbf{U} = \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1/2}$ is a unitary matrix and $\mathbf{S}$ is a semi positive definite matrix. It is then shown by analyzing the eigenvalues of the matrix $\mathbf{M}$ that $\mathbf{R} = \mathbf{U}$ is the matrix minimizing the sum of squared errors.

2.2. The UME - Universal Manifold Embedding

In this section we briefly review the principles of the UME [9] for observations related by an affine transformation. Let $\mathcal{O}$ be the space of observations. Let Φ be the group of affine transformations, and let S be a set of known objects. Every observation is the result of applying a geometric deformation $\phi \in \Phi$ to an object $g \in S$. The parameters of the affine transformation completely specify the action of the group of geometric transformations the object may undergo. We denote by $S_g \in \mathcal{O}$ the set of all possible observations on an object g. Thus, S_g is the orbit of g under Φ.

The universal manifold embedding is a map $T \colon \mathcal{O} \to H$ from the space of observations into a low dimensional Euclidean space, H, such that the set $T(S_g)$ is a linear subspace of H for any g. Thus, the UME reduces the dimension of any problem concerning the multiplicity of appearances of objects from the high dimensional space of observations $\mathcal{O}$ to the low dimensional linear space H and allows for the usage of classical linear theory in the face of a highly non-linear problem.

Next, the mapping T is described. Consider the case where $h(\mathbf{x}), g(\mathbf{x})$ are observations on the same object related by an affine transformation, *i.e.*

$$h(\mathbf{x}) = g(\mathbf{A}\mathbf{x} + \mathbf{c})\ , \mathbf{A} \in \mathrm{GL}(n), \mathbf{c}, \mathbf{x} \in R^n \tag{3}$$

Let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{c}$, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} + \mathbf{b}$ where $\mathbf{b} = -\mathbf{A}^{-1}\mathbf{c}$. Let $\mathbf{y}' = [1, y_1, ..., y_n]^T$ then $\mathbf{x} = \mathbf{D}\mathbf{y}'$ where $\mathbf{D} = \left[\mathbf{b}; \mathbf{A}^{-1}\right]$ is an $n \times n+1$ matrix.

Also, let $P \in \mathbb{N}$ and let w_l, $l = 1, \ldots, P$ be a set of bounded, Lebesgue measurable functions $w_l \colon \mathbb{R} \to \mathbb{R}$. By a change of variables we obtain the following identities:

$$\int_{\mathbb{R}^n} w_l \circ h(\mathbf{x}) d\mathbf{x} = \left|\mathbf{A}^{-1}\right| \int_{\mathbb{R}^n} w_l \circ g(\mathbf{y}) d\mathbf{y} \tag{4}$$

$$\int_{\mathbb{R}^n} \mathbf{x} w_l \circ h(\mathbf{x}) d\mathbf{x} = \left|\mathbf{A}^{-1}\right| \int_{\mathbb{R}^n} (\mathbf{D}\mathbf{y}') w_l \circ g(\mathbf{y}') d\mathbf{y} \tag{5}$$

Let f be some observation on a deformable object and let

$$\mathbf{T}(f) = \begin{bmatrix} \int\limits_{\mathbb{R}^n} w_1 \circ f(\mathbf{y}) & \int\limits_{\mathbb{R}^n} y_1 w_1 \circ f(\mathbf{y}) & \ldots & \int\limits_{\mathbb{R}^2} y_n w_1 \circ f(\mathbf{y}) \\ & \vdots & & \\ \int\limits_{\mathbb{R}^n} w_P \circ f(\mathbf{y}) & \int\limits_{\mathbb{R}^n} y_1 w_P \circ f(\mathbf{y}) & \ldots & \int\limits_{\mathbb{R}^2} y_n w_P \circ f(\mathbf{y}) \end{bmatrix} \tag{6}$$

Let $\mathbf{D}' = [\mathbf{e}_1; \mathbf{D}^T]$, where $\mathbf{e}_1 = [1\,0\,...0]^T$, be the matrix representation of an affine transformation in homogeneous coordinates. Rewriting (4), (5) for $l = 1, \ldots, P$ in a matrix form, we have:

$$\mathbf{T}(g)\mathbf{D}' \left|\mathbf{A}^{-1}\right| = \mathbf{T}(h)\ . \tag{7}$$

To find the matrix $\mathbf{D}'$ (and thus recover the parameters of the affine transformation) we notice that (7) is in fact an over determined linear equation system. Hence, the least squares solution for $\mathbf{D}'$ is given by

$$\mathbf{D}' = |\mathbf{A}| \left[\mathbf{T}(g)^T\mathbf{T}(g)\right]^{-1}\mathbf{T}(g)^T\mathbf{T}(h) \tag{8}$$

3. THE UME FOR RIGID TRANSFORMATIONS

In the following, we assume that $\mathcal{O}$ is the space of all n-dimensional functions from $\mathbb{R}^n$ to $\mathbb{R}$. As an example one may consider a 3D topographical models where the 3D surface is described by specifying the height of each point, *i.e.* , as a function from $\mathbb{R}^2$ to $\mathbb{R}$. Assuming further that the objects undergo rigid geometric transformations of coordinates, $\mathbf{\Phi}$ becomes the special Euclidean group in n dimensions, $SE(n)$. (We note that in Section 4, we demonstrate the applicability of the proposed solution for the registration of such 3D surfaces undergoing rigid geometric deformations).

Since $SE(n)$ is a subgroup of the affine group, the mapping $T\colon \mathcal{O} \to H$ in (6) is still valid. The coordinate transformation however is now different. As the geometric transformation is assumed to be rigid, the matrix $\mathbf{A}$ becomes an orthonormal matrix with determinant of 1, *i.e.* a rotation matrix. Let us denote this rotation matrix by $\mathbf{R}$. In this special case of a rigid deformation, $\mathbf{D}'$ will be denoted by $\mathbf{D}'(\mathbf{R},\mathbf{c})$ to specify the transformation parameters and to distinguish the notation from the one used in the case of affine transformations. Thus (7) is rewritten in the following form:

$$\mathbf{T}(g)\mathbf{D}'(\mathbf{R},\mathbf{c}) = \mathbf{T}(h) \tag{9}$$

Thus, (9) establishes the basic relation of the Rigid Transformation UME (RT-UME). Clearly, a solution for $\mathbf{D}'(\mathbf{R},\mathbf{c})$ cannot be found by applying the least squares solution given in (8), as there is no restriction on the solution to provide a rigid transformation. We next show that by enforcing the rigid motion constraints in formalizing the least squares estimate of $\mathbf{D}'(\mathbf{R},\mathbf{c})$ in (8), the solution is obtained along similar lines to the ones employed in [10]:

Define the error matrix $\mathbf{E} = \mathbf{T}(g)\mathbf{D}'(\mathbf{R},\mathbf{c}) - \mathbf{T}(h)$. Let $\mathbf{E}_i$ denote the i-th row of $\mathbf{E}$. Thus,

$$\mathbf{E}_i = \begin{bmatrix} \int\limits_{\mathbb{R}^n} w_i \circ h(\mathbf{x}) - \int\limits_{\mathbb{R}^n} w_i \circ g(\mathbf{y}) \\ \int\limits_{\mathbb{R}^n}\mathbf{x} w_i \circ h(\mathbf{x}) - \mathbf{R}^T\int\limits_{\mathbb{R}^n}\mathbf{y} w_i \circ g(\mathbf{y}) - \mathbf{b}\int\limits_{\mathbb{R}^n} w_i \circ g(\mathbf{y}) \end{bmatrix}^T \tag{10}$$

and

$$||\mathbf{E}_i||^2 = \left\| \int\limits_{\mathbb{R}^n} w_i \circ h(\mathbf{x}) - \int\limits_{\mathbb{R}^n} w_i \circ g(\mathbf{y}) \right\|^2 + \\ + \left\| \int\limits_{\mathbb{R}^n} \mathbf{x} w_i \circ h(\mathbf{x}) - \mathbf{R}^T \int\limits_{\mathbb{R}^n} \mathbf{y} w_i \circ g(\mathbf{y}) - \mathbf{b} \int\limits_{\mathbb{R}^2} w_i \circ g(\mathbf{y}) \right\|^2 \tag{11}$$

The first term is independent of the deformation parameters, therefore it is enough to minimize the sum over the second term. Define $\mathbf{g}_i = \int\limits_{\mathbb{R}^n} \mathbf{y}^T w_i \circ g(\mathbf{y})$, $\mathbf{h}_i = \int\limits_{\mathbb{R}^n} \mathbf{x}^T w_i \circ h(\mathbf{x})$ and $m_i = \int\limits_{\mathbb{R}^n} w_i \circ g(\mathbf{y})$. Minimizing the sum of squared errors (11) for all $i = 1, \ldots, P$ is equivalent to minimizing

$$\sum_{i=1}^{P} \left\| \mathbf{h}_i - \mathbf{g}_i \mathbf{R} - \mathbf{b}^T m_i \right\|^2 \tag{12}$$

This minimization is the same as the minimization problem (2) solved in [10], with the exception that the translation term is multiplied by the mass coefficients m_i. The property of the translation *being identical for all* i *in* (2) is critical in the procedure of solving (2) in [10]. However, following [8], (12) can be rewritten as a weighted sum of squared errors,

$$\sum_{i=1}^{P} m_i^2 \left\| \frac{\mathbf{h}_i}{m_i} - \frac{\mathbf{g}_i}{m_i}\mathbf{R} - \mathbf{b}^T \right\|^2 \tag{13}$$

which is then solved for the desired $\mathbf{R}$ and $\mathbf{b}$ using the procedure described next.

3.1. Finding the Translation

Define the weighted centroid of $\mathbf{h}_i$ and $\mathbf{g}_i$ *i.e.*

$$\mathbf{h}'_i = \frac{\mathbf{h}_i}{m_i} - \frac{\sum_{j=1}^{P} m_j \mathbf{h}_j}{\sum_{j=1}^{P} m_j^2}\,, \quad \mathbf{g}'_i = \frac{\mathbf{g}_i}{m_i} - \frac{\sum_{j=1}^{P} m_j \mathbf{g}_j}{\sum_{j=1}^{P} m_j^2} \tag{14}$$

In addition, define

$$\mathbf{b}' = \mathbf{b}^T - \frac{\sum_{j=1}^{P} m_j \mathbf{h}_j}{\sum_{j=1}^{P} m_j^2} + \frac{\sum_{j=1}^{P} m_j \mathbf{g}_j}{\sum_{j=1}^{P} m_j^2}\mathbf{R} \tag{15}$$

Using these definitions, the sum in (13) can be written as

$$\sum_{i=1}^{P} m_i^2 \left\| \mathbf{h}'_i - \mathbf{g}'_i \mathbf{R} - \mathbf{b}' \right\|^2 \tag{16}$$

Following similar lines to [10] it can be shown that $\mathbf{b}' = 0$ minimizes (16). Therefore, substituting $\mathbf{b}' = 0$ into (15) we have

$$\mathbf{b} = \left[\frac{\sum_{j=1}^{P} m_j \mathbf{h}_j}{\sum_{j=1}^{P} m_j^2} - \frac{\sum_{j=1}^{P} m_j \mathbf{g}_j}{\sum_{j=1}^{P} m_j^2}\mathbf{R} \right]^T \tag{17}$$

3.2. Estimating the Rotation

In order for $\mathbf{R}$ to minimize (13) , it is required to maximize

$$\sum_{i=1}^{P} m_i^2 \mathbf{h}'_i \mathbf{R}^T \mathbf{g}'^T_i = \mathrm{tr}\{\mathbf{R} \sum_{i=1}^{P} m_i^2 \mathbf{h}'^T_i \mathbf{g}'_i\} \tag{18}$$

Define $\mathbf{M} = \sum\limits_{i=1}^{P} m_i^2 \mathbf{h}'^T_i \mathbf{g}'_i$. Rewriting the definition of $\mathbf{M}$ in a matrix form, we have: $\mathbf{M} = \mathbf{H}^T\mathbf{W}\mathbf{G}$ where $\mathbf{W}$ is a $P \times P$ diagonal weight matrix with the i-th term on the diagonal being m_i^2. Note that $\mathbf{M}$ is a square matrix. Since every square matrix can be decomposed to a positive-semi definite matrix $\mathbf{S}$ and a unitary matrix U such that $\mathbf{M} = \mathbf{U}\mathbf{S}$ where

$$\mathbf{S} = (\mathbf{M}\mathbf{M}^T)^{1/2},\ \mathbf{U} = \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1/2} \tag{19}$$

it follows that $\mathbf{R} = \mathbf{U}^T$ is the matrix maximizing $\mathrm{tr}\{\mathbf{R}\mathbf{M}\}$ and thus

$$\mathbf{R} = (\mathbf{M}^T\mathbf{M})^{-1/2}\mathbf{M}^T \tag{20}$$

is the sought after rotation matrix.

4. EXPERIMENTAL RESULTS

In this section we demonstrate the applicability of the registration method derived using the RT-UME to the problem of registration of 3-D point clouds, representing 3-D surfaces having only partial overlap, where the overlap regions differ by unknown rigid transformation and sampling rates. In this example the point clouds are actual 3-D terrain models.

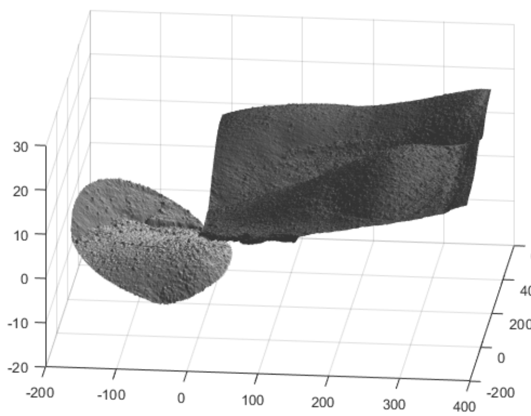


Fig. 1. Example of reference 3-D terrain model and a generated observation

In the presented experiment the RT-UME is used to accurately and efficiently register a terrain patch model (observation) to a larger area 3-D point cloud model (see, Fig. 1) by *jointly* segmenting the corresponding patch out of the larger 3-D reference model and aligning the observation to this patch such that (13) is minimized. The localization is done in two consecutive steps: First a search is conducted over a rough grid, and then over a finer grid in proximity to the minimum point of the rougher grid search. In this experiment the transformation the observation undergoes is a 3-D rigid transformation. The problem is reduced to a 2D problem by attaching local reference axes to the observation and the patch at each point of the grid search. This was done by using principal component analysis to find the axis with least variability in the observation and reference independently and treating it as the Z axis at both. In Fig. 2 the MSE surface is shown over the search grid. It is shown that its minimum is obtained at the correct point (marked by ”x”).

In addition, the performance of the proposed RT-UME in estimating the rigid transformation parameters, is compared to that of the ICP algorithm. In this experiment the ICP is initialized with the ground truth transformation, but as can be concluded from Fig. 3 (a) and Fig. 3 (b) it yields larger estimation errors than the RT-UME. It is therefore concluded that in comparison with the ICP algorithm, the RT-UME provides higher accuracy at a considerably lower computational complexity (as the UME operator is of a linear computational complexity).

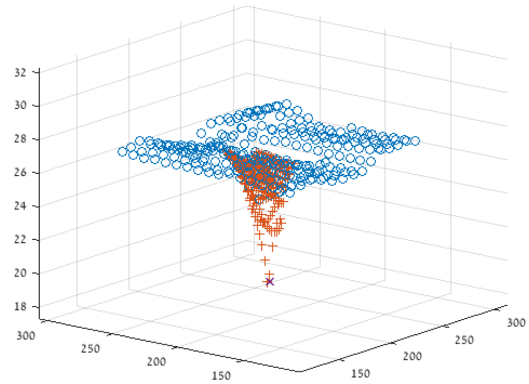
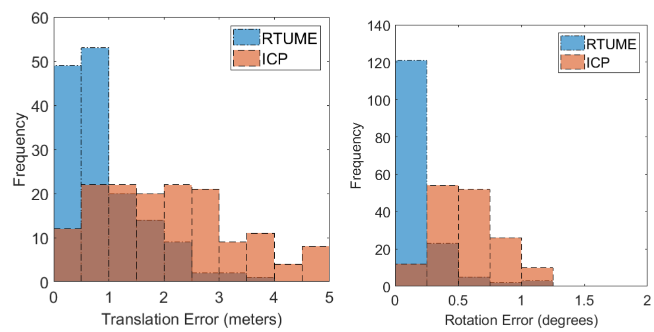


Fig. 2. MSE surface: ”o” depict the results for the rough grid; ”+” depict the results for the second stage fine grid. The correct position of the patch within the larger reference model is marked by ”x”.



(a) (b)

Fig. 3. (a) Translation error distribution of the ICP vs. the RT-UME. (b) Rotation error (solid angle error) distribution of the ICP vs. the RT-UME

5. CONCLUSIONS

We have presented a closed form solution to the problem of registration and detection of dense 3-D point clouds undergoing unknown rigid deformations. The solution is obtained by adapting the general framework of the universal manifold embedding (UME) to the case where the deformations the object may undergo are known to be rigid. Thus, the proposed solution does not require the knowledge of point correspondences, nor an accurate initial registration. We have further demonstrated the applicability of the methodology for the task of registering 3-D point clouds, showing better performance than the commonly used ICP algorithm.

6. REFERENCES

[1] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense Tracking and Mapping in Real-Time," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 2320–2327.

[2] J. Stühmer, S. Gumhold, and D. Cremers, "Real-Time Dense Geometry from a Handheld Camera," in *Pattern Recognition*, M. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, Eds., Berlin, Heidelberg, 2010, pp. 11–20, Springer Berlin Heidelberg.

[3] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multiview Stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, Aug 2010.

[4] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb 1992.

[5] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 119–152, Oct 1994.

[6] H. Pottmann, S. Leopoldseder, and M. Hofer, "Registration without ICP," *Computer Vision and Image Understanding*, vol. 95, no. 1, pp. 54 – 71, 2004.

[7] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan Registration for Autonomous Mining Vehicles Using 3D-NDT," *Journal of Field Robotics*, vol. 24, no. 10, pp. 803–827.

[8] B. K. P. Horn, "Closed-Form Solution of Absolute Orientation Using Unit Quaternions," *J. Opt. Soc. Am.* , vol. 4, no. 4, pp. 629–642, Apr 1987.

[9] R. Hagege and J. M. Francos, "Parametric estimation of affine transformations: An exact linear solution," *Journal of Mathematical Imaging and Vision*, vol. 37, no. 1, pp. 1–16, May 2010.

[10] B. K. P. Horn, H.M. Hilden, and S. Negahdaripour, "Closed-Form Solution of Absolute Orientation Using Orthonormal Matrices," *J. Opt. Soc. Am.*, vol. 5, no. 7, pp. 1127–1135, 1988.