Registration of Noisy Point Clouds using Virtual Interest Points

Mirza Tahir Ahmed*, Mustafa Mohamad†, Joshua A. Marshall*, and Michael Greenspan*

Department of Electrical and Computer Engineering*
School of Computing†
Queen’s University, Kingston ON, Canada
Email: {81mta, 8mm104, joshua.marshall, michael.greenspan}@queensu.ca

Abstract—A new method is presented for robustly and efficiently registering two noisy point clouds. The registration is driven by establishing correspondences of virtual interest points, which do not exist in the original point cloud data, and are defined by the intersection of parametric surfaces extracted from the data. Parametric surfaces, such as planes, exist in abundance in both natural and artificial scenes, and can lead to regions in the data of relatively low noise. This in turn leads to repeatable virtual interest points, with stable locations across overlapping images. Experiments were run using virtual interest points defined by the intersection of three implicit planes, applied to data sets of four environments comprising 100 point clouds. The proposed method outperformed the Iterative Closest Point, Generalized Iterative Closest Point, and a 2.5D SIFT-based RANSAC method in registering overlapping images with a higher success rate, and more efficiently.

Keywords—registration; point cloud; Virtual Interest Point; Key Point; Parametric Surfaces; Planes;

I. INTRODUCTION

Registration is the process of rigidly transforming the coordinate reference frames of two or more 3D point clouds, such that their intersecting regions overlap correctly. Therefore, detection of the area of overlap between two point clouds plays an important role. This is done by finding corresponding points between the two point clouds, using one of two approaches. The first approach assumes that the point clouds are initially approximately aligned, and applies a nearest neighbor search to establish point correspondences based on their proximity in 3D space. The registration process is completed when a convergence criterion is satisfied, such as a minimal threshold of the average separation between correspondences, or a maximum threshold on the number of iterations. Such techniques can be termed non-feature based registration. Iterative Closest Point (ICP) [1] methods [2] are based on this approach.

The second approach is feature based registration, in which the data sets are processed to extract interest points and feature descriptors. In this approach, there is no requirement for an initial approximate alignment of the data sets, and the correspondences are established by determining matches in descriptor space [3], [4]. A transformation is calculated using the correspondences. Typically, this process is wrapped in a robust framework such as RANSAC, to accommodate any false correspondences (i.e., outliers) that result from occasional and inevitable mismatches in descriptor space.

The existence of noise in point cloud data is one source of outliers, and presents challenges in determining correspondences for both feature based and non-feature based registration. Noise is present in the data for several reasons, such as the fundamental physics and measurement technique upon which the sensor is based, which is referred to collectively as sensor-specific noise [5]. Sensor-specific noise includes pixel position error, axial noise, and quantization error. Alternately, scene-specific noise is a result of the sensor’s limitation in correctly observing certain areas in the scene such as corners, edges, and reflective objects [6]. Feature based registration techniques generally identify interest points in regions with large shape variations. Unfortunately, these regions of large variation are among those which are the most highly affected by scene-specific noise, thereby increasing the chances of feature mismatches and registration inaccuracies and failures.

In this paper a novel method is proposed for registering two noisy point clouds. The main idea is to observe the underlying structure of the scene, extract implicit parametric surfaces (such as planes), identify the intersection of multiple parametric surfaces as virtual interest points, and then use these virtual interest points to register the original point clouds. Parametric surfaces such as planes are the most stable areas in the scene because they contain minimal shape variation, and are therefore less susceptible to both sensor-specific and scene-specific noise [5]. The stability of the underlying parametric surfaces can result in stable and repeatable virtual interest points, which leads to accurate registration results. In addition, only one correct correspondence of virtual interest points is sufficient to register two point clouds, which significantly improves the computational efficiency.

A. Related Work

The most widely-used approach for non-feature based registration of two point clouds is the ICP algorithm [1], and many variants of ICP have been developed that use different error functions and point selection strategies. The three most commonly used distance functions are point-to-point, point-to-plane, and plane-to-plane [1], [7], [8]. The transformation
is iteratively refined by applying the computed transformation and recalculating the correspondences. For large initial offsets between two noisy point clouds, however, ICP based approaches will fail to converge [8].

For feature based registration, there exist a number of different 3D interest point operators that are based on different properties of the data, and yet extracting interest points repeatedly among multiple point clouds is a challenge. The Local Surface Patch (LSP) approach [9] searches for areas with large shape variations as measured by shape indices, which are calculated through principal curvature. Intrinsic Shape Signatures (ISS) [10] addresses the detection of viewpoint invariant interest points. 2.5D SIFT [11] detects interest points using an enhanced version of the Lowe’s Scale Invariant Feature Transform (SIFT) algorithm [12]. A discrete scale-space representation by using Gaussian smoothing and Difference Of Gaussian (DOG) pyramid techniques is first created, and maxima are detected within the DOG scale-space. The interest points are finally validated and located by comparing the ratio of the principal curvatures with a predefined threshold.

Once extracted, the support region for each interest point is used to compute a unique feature descriptor. Again there are a number of methods for feature descriptor computation. Signature based methods describe the local neighborhood of an interest point by encoding one or more geometric measures computed individually at each point of a subset of the neighborhood [13]. Histogram based methods describe the local neighborhood of an interest point by accumulating geometric or topological measurements into histograms [14]. A comprehensive overview of a number of interest point and feature descriptor techniques is given in [15].

Given two sets of feature descriptors from two acquired scans, correspondences are computed to find overlapping parts in the data. For matching in feature descriptor space, brute force matching and $k$-d tree nearest-neighbour search (e.g., FLANN) have been applied [16]. Since incorrect correspondences can negatively affect the estimation of the final transformation, some outlier rejection method such as RANSAC is required. The last step is to compute the transformation based on the best correspondences, typically minimizing some least square measure of error.

The proposed method falls under the category of feature based registration. Virtual interest points (VIPs) are identified by observing the underlying structure of the scene. The use of the term virtual highlights that these points do not exist in the original point cloud data; rather, they are 3D points that are injected at distinct locations based on a computation. The VIPs between two point clouds are further annotated, based on the properties of their supporting implicit parametric surfaces, and are matched in feature descriptor space. A rigid transformation to register two point clouds can be computed from a single true VIP correspondence. As discussed in Section III, experimentation has shown improvements in both registration success and computational efficiency over ICP [1], GICP [7], and conventional feature based registration techniques [11].

The remainder of this paper is organized as follows: In Section II, the plane extraction technique is described, as is the criteria to compute VIPs, and the steps of registration using VIPs. In Section III, experimental evidence demonstrates that the method registers noisy point clouds accurately and efficiently compared to other well known techniques. Finally, Section V concludes the paper with a summary and discussion of future work.

II. VIRTUAL INTEREST POINTS FOR REGISTRATION

Repeatable interest point detection presents a challenge in feature based registration due to the presence of noise in the data. To address this, the proposed method limits the search space to detect interest points by avoiding noisy areas where there are large shape variations. Specifically, planar parametric surfaces that contain minimal shape variations are extracted in the point cloud data. Planar regions are less affected by the sensor-specific and scene-specific noise, and are thus considered as the most stable regions [17]. Once such planar surfaces are extracted, the intersection of three non-parallel parametric surfaces result in VIPs.

A. Plane Detection

There are multiple plane extraction algorithms, available such as RANSAC [18], Hough transform [19], and region growing segmentation [20]. We used region growing segmentation. It is based on characterizing the surface curvature of the points in the point cloud. Neighbouring points are
merged together as a region if they are similar enough under the smoothness constraints.

First, all points are sorted by their curvature values so that the region begins its growth from the point that has the minimum curvature value. The surface curvature $\gamma$ for a point $p$ is computed as

$$\gamma = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2},$$

where $\lambda_i$ are the eigenvalues of the covariance matrix for a support region of 50 nearest points in the neighbourhood of $p$. The point with the minimum curvature is identified as residing in a flat (i.e., planar) region, and is used as a seed point.

For every seed point, the algorithm finds the support region around the seed. The following steps are repeated until all the points in point cloud are tested.

1) Every neighbour point is tested for the angle between its normal and normal of the current seed point. If the angle is less than 5 degrees then the current point is added to the current region.

2) If the surface curvature $\gamma$ of the neighbour point is less than a threshold, it is added to the set of seeds. This helps to grow the region beyond the current neighbourhood of the seed point.

3) When all the neighbours of the current seed point are tested, the seed point is removed from the set of seeds.

The output of the algorithm is a set of regions, as shown in Fig. 1 (a) and (b). Seed points are shown in different colors, depicting each planar region. After region growing, points with large curvature values are rejected. Each region is processed further to find out the plane coefficients using eigenvalue decomposition.

Once the plane coefficients are computed, the planes database is populated with the following information:

- **Group index**: Parallel planes are grouped together so that spurious calculations can be avoided.
- **Plane coefficients**: Coefficients of the plane equations, that are used to compute the VIPs.
- **Number of points $P_s$**: A trust measure, such that a greater value of $P_s$ indicates a more reliable plane.

**B. VIPs and Feature Descriptors**

The plane database is processed next to extract VIPs, as shown in Fig. 1 with brown colored blocks. Three planes, each with a different group index (and therefore non-parallel), are processed to find a common point of intersection. The following system of equations of the planes are established to calculate the degree of orthogonality of three planes, and their intersection points

$$A_1x + B_1y + C_1z + D_1 = 0$$

$$A_2x + B_2y + C_2z + D_2 = 0$$

The following two matrices are derived from the above system of planar equations

$$M_c = \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{bmatrix}$$

$$M_a = \begin{bmatrix} A_1 & B_1 & D_1 \\ A_2 & B_2 & D_2 \\ A_3 & B_3 & D_3 \end{bmatrix}$$

where $M_c$ and $M_a$ are the coefficient matrix and the augmented matrix, respectively. Three non-parallel planes can only intersect at a point if both $M_c$ and $M_a$ are full rank. $M_c$ contains the row unit normals for each plane, and the closer the determinant of $M_c$ is to 1, the closer the three planes are to being mutually perpendicular.

If $M_c$ and $M_a$ are full rank, then the point of common intersection, which is a candidate VIP, is calculated using (2) as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{bmatrix}^{-1} \begin{bmatrix} -D_1 \\ -D_2 \\ -D_3 \end{bmatrix}$$

In total, each set of three planes has to satisfy two conditions to generate a qualified VIPs. First, the rank of $M_c$ and $M_a$ must be 3. Second, $\det(M_c) > \sigma$, where $\sigma$ is a threshold that is close to 1 and indicates the degree of orthogonality of the three planes.

Qualified VIPs are added to a database along with their associated feature descriptors, which are composed of the following properties:

1) Angles between the three planes: $\alpha, \beta, \gamma$.
2) Degree of orthogonality: $\det(M_c)$.
3) Trust factor: $t_f$, which is calculated as

$$t_f = \frac{P_{s_1} + P_{s_2} + P_{s_3}}{T_s}$$

where $P_{s_1}$, $P_{s_2}$, and $P_{s_3}$ are the number of points of three associated planes and $T_s$ is the size of point cloud. The higher the value of $t_f$, the more reliable the VIP.

**C. VIP Correspondences**

There a number of methods available for feature matching to establish a set of VIP correspondences. A nearest-neighbour search mechanism is employed to match the VIPs in two point clouds. The $k$-d tree is populated where each property of the feature descriptor serves as a dimension. A source VIP is matched with the nearest target VIP in descriptor space. An obvious selection of the best correspondence is the one with minimum distance in feature space. We only need one VIP true correspondence to register two point clouds. However, it is possible that a selection based on only minimum distance in descriptor space might be an
rotation in one axis is given by

\[ k \]

where \( k \) is a cross product matrix of \( v_1 \times v_2 \).

The translation \( T \) is calculated trivially as the difference between the two VIPs; i.e., \( T = v_{q_1} - v_{m_2} \). To calculate the two rotations, first it is recognized that the two VIPs have three matched planes, so three angles between the matched planes are computed as

\[ \alpha_i = \cos^{-1}\left( \frac{\langle n_{q_i}, n_{m_i} \rangle}{\sqrt{\langle n_{q_i}, n_{q_i} \rangle \cdot \langle n_{m_i}, n_{m_i} \rangle}} \right) \]

where \( \alpha_i \) is the angle between the \( i \)-th normals \( n_{q_i} \) and \( n_{m_i} \) of the two matched VIPs.

We select the two largest angles and calculate the rotation matrix \( R \) using Rodrigues’ rotation formula, in which rotation in one axis is given by

\[ R = I + \sin \alpha_i [k]_x + (1 - \cos \alpha_i)(kk^T - I) \]

where \( k = (n_{q_i} \times n_{m_i}) \) is the cross product of two normals and \( \langle k \rangle_x \) is a cross product matrix of \( k \).

Two point clouds are registered together using the translation matrix \( T \) and the rotation matrix \( R \). Hence using only one true correspondence, two point clouds are registered. This significantly improves the computational efficiency of the registration process.

III. EXPERIMENTAL RESULTS

Four point cloud data sets were collected, using different sensors, comprising of both artificial and natural structures to compare the proposed method against ICP, GICP, and 2.5D SIFT for registration. The sensors used to acquire data were

- Panel is captured by the Asus Xtion. The sensor is moved in half a meter height along a zig-zag structure built from wooden panels. It consists of 15 frames. This data set is available online [21]. The ground truth of this data set is generated using ICP with 500 iterations and verified through visual inspection. ICP registers frames very accurately because of two reasons. First, the
frame offset is very small and secondly, this data set has low noise due to short distance between the sensor and the surface.

- **Lab** consists of 28 consecutive scans of a small area of a laboratory environment, using the Kinect. This data set was captured at a very small distance of average 1.5 m to reduce noise effects. The translation between subsequent scans is very small and contains a large area of overlap.

- **Office** consists of 12 scans captured of an office environment by rotating the Kinect sensor approximately around its vertical axis through a 360° arc. This data set was captured at a distance of on average 3 m to 5 m, and so it contains a higher amount of noise and quantization error than the Lab data set. The interscan displacement is also larger than that of the Lab data set, but still contains a reasonable area of overlap between successive scans.

- **Rocks** represents a natural structure. It consists of 50 consecutive scans of a rock cut, captured near a highway overpass using an SR4030 sensor. This data set was captured at a very small distance of on average 1 m. The observed rock cut was composed of an exposed granite outcrop, which is part of the Canadian shield formation. During the construction of highways, these rocks are blasted and thus the rock faces are revealed to exhibit their underlying semi-regular approximately planar surfaces.

VIP with and without ICP refinement is compared to the ground truth Panel data set. In the case of VIP refined with ICP (VIP-ICP), the iterations of ICP are set to 5 so that the computation efficiency of registration does not degrade. The proposed technique converged to a global minimum for 93% of frame pairs. Camera origin and orientation for 15 frames are tracked with VIP, VIP-ICP, and ground truth. Thus, ground truth is compared to VIP and VIP-ICP for registration performed with only a single correspondence. The success of each method was evaluated qualitatively by visually comparing the convergence to a global minimum. The efficiency of each method was compared on an Intel Core i7 machine with 6 GB of RAM.

On the Lab data set using Kinect, ICP and GICP converged to a global minimum for 86.2 % and 89.6 % of image pairs, respectively. However, 2.5D SIFT converged only for 18.5 % of pairs. The 2.5D SIFT interest points

\[
\text{err}_{\text{PC}} = \frac{T_{\text{err}}}{\text{dia}(PC)} \times 100
\]

where \(T_{\text{err}}\) is the translation error relative to the ground truth and \(\text{dia}(PC)\) is the average diameter of point clouds, \(PC\), of the data set. The percentage error for translation for VIP and VIP-ICP are \((1.76, 0.55, 0.46)\) and \((1.00, 0.28, 0.16)\), respectively. The average rotational error in degrees is computed as \((0.64, 1.76, 1.14)\) and \((0.69, 0.74, 0.81)\) for VIP and VIP-ICP, respectively. It can be seen that the error in each of the 6DoF is very low for both VIP and VIP-ICP, thus the quality of registration is good. It can be visually inspected in Fig. 2. Similarly, there is not a significant improvement if ICP refinement is added as a post processing step to the proposed technique. However, the average computation time per pair for VIP-ICP, 4.27 s, is almost double than VIP, 2.41 s. Therefore, VIP is used for registration for the rest of the data sets.

Registration applied to three data sets using four different techniques. Two variants of iterative closest point algorithm, ICP and GICP, were compared as non-feature based registration methods. The number of iterations are limited to a maximum of 100 since convergence is typically achieved before this point, if at all. The rate of convergence predictably dropped by reducing the number of iterations in ICP and GICP. 2.5D SIFT is used for feature based registration, as is the VIP method described above. The feature descriptor, Fast Point Feature Histograms (FPFH) [14], is used for feature matching of the extracted 2.5D SIFT interest points. The success of each method was evaluated qualitatively by visually comparing the convergence to a global minimum. The efficiency of each method was compared on an Intel Core i7 machine with 6 GB of RAM.

![Figure 3. Lab data set registration of two point clouds. (a) Two raw point clouds are displayed from a bird’s eye view without registration to illustrate the displacement between them. (b) Registration using ICP method that takes 100 iteration to converge to a global minimum. (c) GICP registers better with 100 iteration but still not accurate . (d) 2.5D SIFT fails to align two point clouds. (e) VIP registers better than other techniques, with registration performed with only a single correspondence.](image-url)
were not highly repeatable because it searched for interest points at regions of large shape variation that suffer due to scene-specific noise. For this reason, the registration failed even with significant overlap between the point clouds. On the other hand, VIP successfully registered all 27 pairs of images (100%). Fig. 3 illustrates the registration results using the four techniques on a pair of scans. The average processing time of VIP is 2.81 s per pair. 2.5D SIFT has an average processing time of 25.0 s per pair. ICP with 77.1 s per pair is slightly more efficient than GICP with 171.3 s per pair. Therefore, VIP significantly outperformed the other three techniques in computational efficiency.

On the Office data set using the Kinect sensor, ICP and GICP converged to a global minimum for 40.9 % and 27.3 % of image pairs out of 11 pairs, respectively. However, 2.5D SIFT converged for only 0.09 % of pairs. On the other hand, VIP successfully converged for 90.9 % of 11 pairs, failing where the plane detection was erroneous. This is a noisy data set due to the quantization error that resulted from the acquisition procedure. Fig. 5 illustrates the registration using the four techniques on a pair of scans. 2.5D SIFT converged only for a single pair with a runtime of 45.3 s per pair. The computation time for ICP and GICP was 169.91 and 567.45 s per pair, respectively. ICP and GICP converged very slowly due to the large scan to scan displacement. Whereas VIP registered efficiently with an average processing time of 0.6 s per pair.

The Rocks data set was acquired using SR4030 sensor. Since only depth information is extracted in this data set, a RGB image of a rock is shown for illustration in Fig. 4(a). This data set contained a large amount of noise and due to the lack of color information, it is difficult to distinguish between two point clouds. Therefore, the two clouds are shown in different colors. ICP and GICP converged to a global minimum for 85.7 % and 87.7 % of pairs, respectively. 2.5D SIFT failed to converge due to non-repeatability of interest point because of noise in the data set. It converged for only 0.12 % of pairs. On the other hand, VIP performed well on this data, as it converged for 93.9 % of pairs. The efficiency performance of all the algorithms was better for this data set due to the fewer number of points in the point clouds. VIP shows slight improvement in performance as compared to others because the computation is based on parametric surfaces rather than number of points in the point cloud. The average processing time of VIP is 2.32 s. The results are summarized in Table I.

Registration normally fails due to noise in the data...
Table I

<table>
<thead>
<tr>
<th>Data set</th>
<th>Pairs</th>
<th>ICP</th>
<th>GICP</th>
<th>2.5D SIFT</th>
<th>VIP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Con. (%)</td>
<td>Avg. Time</td>
<td>Itr.</td>
<td>Con. (%)</td>
</tr>
<tr>
<td>Lab</td>
<td>27</td>
<td>86.20</td>
<td>77.13</td>
<td>100</td>
<td>89.05</td>
</tr>
<tr>
<td>Office</td>
<td>11</td>
<td>40.90</td>
<td>169.91</td>
<td>100</td>
<td>27.27</td>
</tr>
<tr>
<td>Rocks</td>
<td>49</td>
<td>85.71</td>
<td>7.84</td>
<td>100</td>
<td>87.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.81</td>
<td>93.87</td>
<td>2.32</td>
<td>90.9</td>
</tr>
</tbody>
</table>

Figure 6. Convergence degradation due to noise. Noise standard deviations \( \sigma = 0.0001, 0.001, \) and 0.01 are added to two point clouds and tested with ICP, 2.5D SIFT, and VIP.

sets. Another experiment was performed to test the convergences in the presence of noise. We added Gaussian white noise with zero mean and standard deviation \( \sigma = 0.0001, 0.001, 0.01 \) m to the two point clouds shown in Fig. 1 and performed registration using ICP, 2.5D SIFT, and VIP. The results are given in Fig. 6. It can be noticed that the structure is not visible when \( \sigma = 0.01 \). It is noticed that 2.5D SIFT failed to converge even for noise standard deviation of \( \sigma = 0.0001 \) because features are not repeatable in the presence of noise. However, ICP, which is non-feature based registration technique, converged to a global minimum for \( \sigma = 0.0001, 0.001 \). However, an offset between two point clouds can be seen, if observed closely, when the \( \sigma = 0.01 \). On the other hand, our technique, which is feature based, converged to a global minimum for \( \sigma = 0.0001, 0.001, 0.01 \). The reason for the convergence of VIP is that the saliency measure of extracting parametric surfaces is less affected by noise.

IV. Discussion

VIP outperformed all other tested techniques in terms of convergence and efficiency in all three data sets. In the noisy data sets where measurement noise and quantization error was very high, VIP was the most successful because of its saliency measure of searching for stable areas in the point cloud. Parametric surfaces absorb the affect of measurement noise and quantization error. Furthermore VIPs are not based on a single planar surface but a group of three, which is the reason for the good repeatability of VIPs even in the presence of a significant amount of noise. However, the limitation of VIP is that it requires at least three non-parallel planes in the point cloud to compute a single VIP. In the presence of only parallel planes, it normally fails which is an issue to be addressed in future work. Nevertheless, point clouds normally acquired from natural and artificial environments are rich in areas that exhibit planar surfaces. Using ICP as a post processing step for further refinement improves the quality of registration and may also register only parallel planes but decreases the runtime of the algorithm. Only parallel planes cases may be solved by adopting the planes correspondences determination technique [22].

For ICP and GICP, runtime increases almost linearly with the size of the point cloud, as shown in Table I. Convergence decreases when the displacement is large between the point clouds. On the other hand, interest point detection techniques, such as 2.5D SIFT that use large shape variations as saliency measure, cannot successfully register noisy point clouds because such areas contain significant amount of noise due to the sensors’ limitations.

V. Conclusion

A novel robust feature based registration method is presented. The method is robust to noise because avoiding noisy areas is one of the key aspect considered in design criteria.
The VIPs are computed by observing the most stable areas in the scene, therefore even in the presence of noise the performance of the algorithm does not degrade. Furthermore, the feature descriptors are computed in such a way that only one true correspondence is required for registration of noisy point clouds. As a result, compared to other commonly used registration methods, the proposed method is more stable and computationally efficient. The method was experimentally evaluated and shown to outperform the PCL versions of ICP, GICP, and 2.5D SIFT, in terms of convergence behaviour and efficiency.

In future work, this method will be expanded to solve the aforementioned cases by including other types of parametric surfaces, such as lines, spheres, and conics. The technique will also be tested against other natural scenes, and generalized for object recognition purposes.

REFERENCES


