Fast Multiview 3D Scan Registration using Planar Structures

Uttaran Bhattacharya, Sumit Veerawal, Venu Madhav Govindu*
Indian Institute of Science
Bengaluru 560012 INDIA
uttaran.bhattacharya127@gmail.com | sumit.veeru@gmail.com | venu@ee.iisc.ernet.in

Abstract

We present a fast and lightweight method for 3D registration of scenes by exploiting the presence of planar regions. Since planes can be easily and accurately represented by parametric models, we can both efficiently and accurately solve for the motion between pairs of 3D scans. Additionally, our method can also utilize the available non-planar regions if necessary to resolve motion ambiguities. The result is a fast and accurate method for 3D scan registration that can also be easily utilized in a multiview registration framework based on motion averaging. We present extensive results on datasets containing planar regions to demonstrate that our method yields results comparable in accuracy with the state-of-the-art while only taking a fraction of computation time compared with conventional approaches that are based on motion estimates through 3D point correspondences.

1. Introduction

Modern depth scanners are capable of acquiring good quality scans of 3D environments which can be utilized to generate 3D maps for the purposes of autonomous robotic navigation, area monitoring and many other applications. This growth in 3D scanning capabilities entails a concomitant need for fast processing pipelines capable of generating accurate 3D scene models by registration of individual 3D scans. However, most of the available 3D registration techniques based on point correspondences between scans do not scale well for faster implementations. In the current paper, we exploit the presence of planes in common real world indoor environments to build a registration pipeline that is fast, lightweight and provides sufficient accuracy for practical applications. Our approach first identifies and segments planar regions in the input scans using a simple and efficient clustering scheme that utilizes knowledge of the sensor noise model. Subsequently we match planar regions across scans. This matching of planes is accomplished by utilizing the fact that plane normals are invariant to Euclidean motion and at the same time yield discriminative matching. We use the available 3D planar regions matched across scans to robustly solve for the motion parameters between pairs of scans. In certain cases, when we have very few planes in scans, the planar representations do not yield a sufficient number of constraints for motion estimation. In such cases, we augment the motion estimation procedure by also utilizing a few point correspondences in non-planar regions of the scans. Finally, we use the individual pairwise motion estimates in a robust motion averaging framework to solve for the global multiview registration problem, i.e. place all scans in a common frame of reference.

*Corresponding author

2. Related Work

The key components of our 3D registration pipeline are identification and segmentation of planar regions in 3D scans, matching of such planes across scans, estimation of relative motion between matched planar representations and global multiview registration through motion averaging. While the literature of 3D scan registration is very large, in this Section, we limit ourselves to a brief review of literature relevant to the individual steps in our pipeline.

Lee *et al*. [13] used a RANSAC based approach and uncertainty analysis to perform plane segmentation. The plane-based simultaneous localization and mapping (Planar-SLAM) algorithm of Siegwart *et al*. [23] also uses uncertainty analysis with least-square fitting for extracting planes. Pathak *et al*. [16] solved for plane segmentation of scans collected with extremely noisy range sensors using uncertainty analysis and region growing. Feng *et al*. [5] used agglomerative hierarchical clustering to perform plane segmentation. In comparison with planar segmentation, the problem of plane correspondence across scans has received less attention in the 3D registration literature. Most approaches use heuristics such as color features, area similarity, normal similarity of planar regions etc. However,

Pathak *et al*. [15] solved for the unknown correspondences using a plane-parameter covariance matrix after pruning the search space on the basis of shape factor, area ratio, curvature histogram and inter-surface relations.

For motion estimation based on plane parameters, Siegwart *et al*. [23] used an extended Kalman filter (EKF) approach. Taguchi *et al*. in their point-plane based SLAM approach [20] solved for motions first using a joint plane-point model and then applying a global optimization on all matching 3D feature points and 3D planes. Pathak *et al*. [15] used least-squares estimation of rotation to align plane normals. They also solved for translation using a rough estimate of plane overlap and a small translation approximation. More general motion estimation methods typically use variants of the ICP algorithm [17]. For instance, the very popular KinectFusion [14] method is based on ICP and works well when the depth sensor is moved slowly and smoothly in the scene. KinectFusion registers each scan with an accumulative map it has built based on previous scans. However, it does not take loop closures into account and suffers from drift errors. Kintinuous [24], a modified version of KinectFusion, allows dynamically updating the point cloud data being mapped and resorts to fast odometry from vision on scans where ICP does not converge.

With regard to multiview 3D registration using pairwise motion estimates, Govindu [8] introduced a motion averaging approach that reduced the effects of accumulated drift errors in global registration in a geometrically consistent manner. This approach was made robust by incorporating a RANSAC step that identified and rejected individual motion outliers [9]. More recently, Choi *et al*. [3] developed an indoor scene reconstruction approach using dense and reliable point correspondences and running a robust global optimization with outlier removal. Zhou *et al*. [27] modified this method to drastically increase the speed of the algorithm without significantly compromising accuracy.

3. Problem Formulation

In this Section, we show how we can utilize the representations of available 3D planes to solve for the motion between a pair of 3D scans. Let there be m_k planes in the k-th scan. We represent the set of all planes in the k-th scan as $\mathbf{S}_k = \begin{bmatrix}\mathbf{s}_1 & \mathbf{s}_2 & \dots & \mathbf{s}_{m_k}\end{bmatrix}$ such that for a point $\mathbf{p} = \begin{bmatrix}X & Y & Z\end{bmatrix}^\top$ lying on plane l in the k-th scan, $l \in \{1, \dots, m_k\}$, we have $\mathbf{s}_l^\top \mathbf{p} + 1 = 0$. Note that we can also equivalently represent the plane parameters by the matrix $\mathbf{L}_k = \begin{bmatrix}\mathbf{N}_k^\top & \mathbf{d}_k\end{bmatrix}^\top$, $\mathbf{N}_k = \begin{bmatrix}\mathbf{n}_1 & \mathbf{n}_2 \cdots \mathbf{n}_{m_k}\end{bmatrix}$, $\mathbf{d}_k = \begin{bmatrix}d_1 & d_2 \cdots d_{m_k}\end{bmatrix}^\top$, where $\mathbf{n}_l = \frac{\mathbf{s}_l}{\|\mathbf{s}_l\|_2}$ is the *unit normal* to plane l in the scene and $d_l = \frac{1}{\|\mathbf{s}_l\|_2}$ is the *offset* of the same plane l from the origin of the frame of reference. Thus, for the point $\mathbf{p}$ mentioned above, we can now write,

$$\mathbf{n}_l^\top \mathbf{p} + d_l = 0 \tag{1}$$

$$\implies \mathbf{M}^{-\top} \begin{bmatrix}\mathbf{n}_l & d_l\end{bmatrix}^\top \mathbf{M} \begin{bmatrix}\mathbf{p} \\ 1\end{bmatrix} = 0 \tag{2}$$

where $\mathbf{M} = \begin{bmatrix}\mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^\top & 1\end{bmatrix}$ represents the rigid motion of rotation $\mathbf{R} \in \mathbb{SO}(3)$ followed by translation $\mathbf{t} \in \mathbb{R}^3$. We observe from Equation (2) that when the point $\mathbf{p}$ on the plane is transformed by $\mathbf{M}$ to the point $\mathbf{p}'$ (say), the transformed plane parameter $\begin{bmatrix}\mathbf{n}_l' & d_l'\end{bmatrix}^\top$ are given in terms of the initial plane parameters by

$$\begin{bmatrix}\mathbf{n}_l' \\ d_l'\end{bmatrix} = \mathbf{M}^{-\top} \begin{bmatrix}\mathbf{n}_l \\ d_l\end{bmatrix} \tag{3}$$

$$\implies \begin{bmatrix}\mathbf{n}_l' \\ d_l'\end{bmatrix} = \begin{bmatrix}\mathbf{R}\mathbf{n}_l \\ -\mathbf{t}^\top \mathbf{R}\mathbf{n}_l + d_l\end{bmatrix} \tag{4}$$

Now, given a pair of plane parameter matrices $\mathbf{L}_i$ and $\mathbf{L}_j$ and a set of correspondences between their columns denoted by the set $\mathcal{K}_{ij}$, we aim to estimate $\mathbf{R}$ and $\mathbf{t}$ that satisfies Equation (4), *i.e*., we solve the optimization:

$$\min_{\{\mathbf{R},\mathbf{t}\}} \sum_{\mathcal{K}_{ij}} \phi\left(\begin{bmatrix}\mathbf{n}_{j_l} \\ d_{j_l}\end{bmatrix} - \begin{bmatrix}\mathbf{R}\mathbf{n}_{i_l} \\ -\mathbf{t}^\top \mathbf{R}\mathbf{n}_{i_l} + d_{i_l}\end{bmatrix} \right) \tag{5}$$

where the tuple $(i_l, j_l) \in \mathcal{K}_{ij}$ denotes correspondence indices. Note that the solution of $\mathbf{R}$ in Equation (5) depends only on the corresponded plane normals and is independent of the solution of $\mathbf{t}$. Further, if we choose $\phi(\cdot) = \|\cdot\|_2^2$ as the distance metric, then the optimal least-squares estimate of $\mathbf{R}$ is given by Umeyama's method [22], simplified for a pure rotation scenario. Having estimated $\mathbf{R}$, the optimal least-squares estimate of $\mathbf{t}$ becomes the solution of a system of linear equations in $\mathbf{t}$:

$$\begin{bmatrix}\mathbf{n}_{i_1}^\top \mathbf{R} \\ \mathbf{n}_{i_2}^\top \mathbf{R} \\ \vdots \\ \mathbf{n}_{i_X}^\top \mathbf{R}\end{bmatrix} \mathbf{t} = \begin{bmatrix}d_{i_1} - d_{j_1} \\ d_{i_2} - d_{j_2} \\ \vdots \\ d_{i_X} - d_{j_X}\end{bmatrix} \tag{6}$$

where we must have $X = |\mathcal{K}_{ij}| \geq 3$ for a non-trivial algebraic solution to exist.

4. The Registration Pipeline

In this Section, we list and describe the various components of our pipeline.

1. **Plane Segmentation:** Segment planar regions from raw depth scans by incorporating a sensor noise model,

2. **Plane Correspondence:** Solve for correspondence between the plane representations across scans using invariance to the underlying motion of plane positions relative to each other within scans,

3. **Motion Estimation:** Solve for motions between scans using parameters of corresponded planes; also use non-planar regions in a robust modification of Umeyama's method [22] if required,

4. **Global Registration:** Register all scans to a global frame of reference using robust motion averaging of the estimated pairwise motions.

We detail each of these steps in turn.

4.1. Plane Segmentation

It is well-known that for stereo based structured-light depth sensors, the noise or uncertainty in depth varies quadratically with depth. However, for the purposes of segmentation, we use the disparity map as a more convenient representation since disparity has uniform uncertainty that is independent of the disparity value. In our approach in this subsection, we closely follow the proposal of [2]. Suppose a 3D point $\begin{bmatrix} X & Y & Z \end{bmatrix}^\top$ is projected to a pixel location $\begin{bmatrix} x & y \end{bmatrix}^\top$ in the sensor. Then we have,

$$x = \frac{fX}{Z} + u; \quad y = \frac{fY}{Z} + v \tag{7}$$

where f is the focal length and (u, v) is the principal point of the sensor. The disparity at point $\begin{bmatrix} x & y \end{bmatrix}^\top$ is given by $D(x, y) = \frac{fB}{Z}$, where B is the baseline distance between the projector and the camera center. Now if $\begin{bmatrix} X & Y & Z \end{bmatrix}^\top$ lies on a 3D plane satisfying the equation: $aX + bY + cZ + 1 = 0$, then by multiplying both sides by $\frac{f}{Z}$, we have

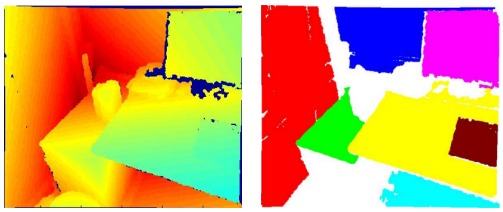
$$a(x - u) + b(y - v) + cf + \frac{f}{Z} = 0 \tag{8}$$

$$\implies ax + by + \frac{D(x, y)}{B} + (cf - au - bv) = 0 \tag{9}$$

Thus, we have an affine relationship between the image points corresponding to planar regions and their disparity map, which we use to carry out plane segmentation. In practice, we pass each noise-filtered disparity map through a Laplacian-of-Gaussian (LoG) filter, which has a high response for sharp changes in the input and zero response for smooth regions, to detect the planar regions. In other words, up to noise, in planar regions we expect the LoG filter to have a zero response. Once we have the association of points in a disparity map with planes thus detected, we proceed to estimate the plane parameters as follows. Given coplanar points of the form $\begin{bmatrix} x_i & y_i \end{bmatrix}^\top$ corresponding to 3D points $\begin{bmatrix} X_i & Y_i & Z_i \end{bmatrix}^\top$, $i = 1, 2, \ldots, n$, $n \geq 3$, we have from Equation (8),

$$\begin{bmatrix} x_1 - u & y_1 - v & f \\ x_2 - u & y_2 - v & f \\ \vdots & \vdots & \vdots \\ x_n - u & y_n - v & f \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = -f \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix} \tag{10}$$

We solve Equation (10) for $\begin{bmatrix} a & b & c \end{bmatrix}^\top$ using standard least-squares optimization techniques. Next, we perform k-means clustering on the estimated plane parameters to find the largest coherent regions in each map corresponding to 3D planes. Figure 1 shows an example of planes segmented from a raw depth map using this approach.



(a) Raw depth map *(b) Segmented planes*

Figure 1: Segmentation of raw depth maps into planar regions

4.2. Plane Correspondence

Once the scans are segmented into planar regions, we need to find correspondences between planar regions across scans in order to find the relative motions between the scans. Given a pair of plane segmented 3D scans containing m_i and m_j planes respectively, we collect all the corresponding plane parameters from the first scan in the matrix $\mathbf{L}_i$ and similarly from the second scan in a matrix $\mathbf{L}_j$, as detailed in Section 3. The correspondence of planes between the scans, $\mathcal{K}_{ij}$, is then ideally obtained from the result of operating a *permutation matrix* $\mathbf{K}_{ij}$ that takes us from $\mathbf{L}_i$ to $\mathbf{L}_j$ (or vice versa), *i.e.*, in a least-squares optimization sense,

$$\begin{aligned} \mathbf{K}_{ij} = \arg\min_{\mathbf{K}} \|\mathbf{L}_j - \mathbf{L}_i\mathbf{K}\|^2 \\ \textit{s.t. } \mathbf{k}_l \in \{\mathbf{e}_1, \ldots, \mathbf{e}_{m_i}, \mathbf{0}_{m_i}\} \text{ for } l = 1, \ldots, m_i \end{aligned} \tag{11}$$

where $\mathbf{K} \triangleq \begin{bmatrix} \mathbf{k}_1 & \ldots & \mathbf{k}_{m_i} \end{bmatrix}$, we consider $m_i \geq m_j$ and the set $\{\mathbf{e}_1, \ldots, \mathbf{e}_{m_i}\}$ represents the m_i-dimensional canonical bases. Note that some columns of $\mathbf{K}_{ij}$ may have only zeros, indicating that the respective columns in $\mathbf{L}_i$ do not have any correspondence in $\mathbf{L}_j$.
However, the permutation matrix is hard to solve for directly in an optimization setup. There are two common approaches to solve for this permutation matrix analytically.

- Considering 3D points in place of plane normals, we can view the point features (representations of the

points in the scans that take into account the local geometric properties of shape) to be corresponded between the two scans as the disjoint vertex sets of a weighted bipartite graph, and find an injection between the sets with minimum weight. Gelfand *et al*. [6] developed a popular solution to this problem, where they exploit the rigidity of the motion between the scans to prune the search space of a branch and bound algorithm used to obtain the required correspondences, and make the algorithm fast and efficient.

- **Graph Isomorphism approach**, where the points in the two scans form the vertices of two graphs respectively, and the weights of the corresponding edges of the graphs are based on the chosen measure between the intra-scan points that is invariant to the underlying rigid motion. The correspondence problem is then recast as the problem of finding an isomorphism between the graphs of the two scans, which has known efficient analytic solutions [21, 1, 19].

In the present scenario, we have plane parameters in place of actual points. Given the two sets of plane normals in $\mathbf{N}_i$ and $\mathbf{N}_j$, one can easily show that the angles between the plane normals within each scan are invariant to any rigid motion applied on the scans, indicating that the graph isomorphism approach is more suited to our representation. Thus we define fully connected graphs $\mathcal{G}_i = \{\mathcal{V}_i, \mathcal{E}_i, \mathcal{W}_i\}$ and $\mathcal{G}_j = \{\mathcal{V}_j, \mathcal{E}_j, \mathcal{W}_j\}$ respectively for the scans i and j such that:

- $\mathcal{V}_i$ and $\mathcal{V}_j$ represent the columns of the plane normal sets in corresponding scans i and j,
- for all edge in $\mathcal{E}_i$, its weight in $\mathcal{W}_i$ is the angle between the two normals (vertices) it joins, and similarly
- for all edge in $\mathcal{E}_j$, its weight in $\mathcal{W}_j$ is the angle between the two normals (vertices) it joins.

The solution of the permutation matrix $\mathbf{K}$ described in Equation (11) can now be recast as one of finding an isomorphism between the graphs $\mathcal{G}_i$ and $\mathcal{G}_j$. One of the most prominent solutions to solving for the weighted graph isomorphism problem (WGIP) analytically has been given by Umeyama [21], which finds an isomorphism between weighted graphs with equal number of vertices if they are isomorphic or nearly isomorphic. In this paper, we use a popular software implementation of this solution in the GraphM package [26], which pads dummy nodes wherever required without affecting the process of Umeyama's solution, so that arbitrary input graph pairs always have equal number of vertices.

4.2.1 Ill-definition of the WGIP

The WGIP problem becomes ill-defined when

1. all the edges in the graph have the same weight, *e.g*., the corner of a room with two walls and the floor (or the ceiling) that are mutually orthogonal, or

2. the graph consists of only two vertices and one edge connecting them, which occurs frequently in practice when only two planes are detected in scenes, *e.g*., two walls of a room.

In both these cases, all possible permutation matrices produce the same optimization cost, and are thus deemed equally likely to be the correct permutation. We circumvent this issue in practice by assuming a small enough motion between adjacent scans (as is common in datasets used in SLAM, navigation and other such common scenarios) and applying nearest neighbor matching on the columns of their plane parameter sets $\mathbf{L}_i$ and $\mathbf{L}_j$.

We also note that planes that are parallel to each other within scenes have normals that form equivalence classes with respect to any rotation applied on them. Thus, in order to recover a non-trivial rotation between a pair of scans using matched plane normals, we should pick a single member from each equivalence class of normals from either scan and find correspondences only between them. For practical considerations of noise in the estimation of the plane parameters, we pick the plane with the largest number of pixels from each equivalence class, which will have the noise averaged out over more points on the plane than others. Figure 2 demonstrates this procedure for a sample pair of scans, each containing six planes belonging to three equivalence classes.

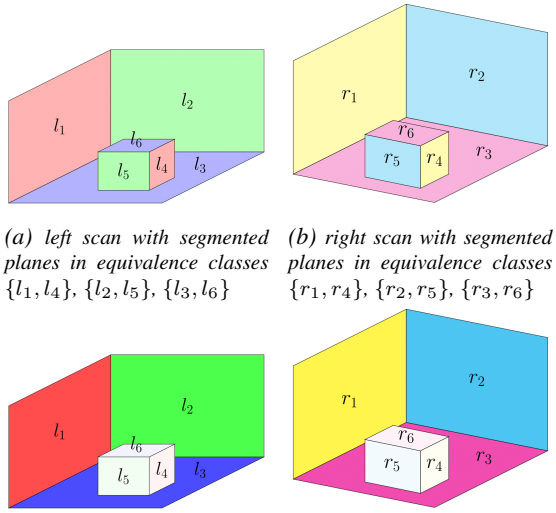
4.3. Motion Estimation

As noted earlier in Section 3, we solve for the rotation first, followed by solving for the translation.

4.3.1 Rotation Estimation

The optimal solution to the underlying rigid motion in the least-squares sense based on point correspondences has been given by Umeyama [22]. In the current approach, we estimate only the optimal rotation with Umeyama's method using the correspondence between plane normals clustered into equivalence classes. Equations (2) and (4) together show that the same rotation that applies between the corresponded points across scans, applies between the corresponded plane normals as well. Thus, if $\mathbf{N}_i$ and $\mathbf{N}_j$ contain the representative plane normals of the equivalence classes in scans i and j respectively, then the rotation $\mathbf{R}_{ij}$ going from scan i to scan j is given by,

$$\mathbf{R}_{ij} = \mathbf{U}\mathbf{S}\mathbf{V}^\top \tag{12}$$



(a) left scan with segmented planes in equivalence classes $\{l_1, l_4\}$, $\{l_2, l_5\}$, $\{l_3, l_6\}$

(b) right scan with segmented planes in equivalence classes $\{r_1, r_4\}$, $\{r_2, r_5\}$, $\{r_3, r_6\}$

(c) left scan with highlighted largest plane from each equivalence class: l_1, l_2, l_3

(d) right scan with highlighted largest plane from each equivalence class: r_1, r_2, r_3

Figure 2: Classification of planes into equivalence classes: (2a) and (2b) show the parallel planes in the left and the right scans have been classified into equivalence classes and are shown in same color. (2c) and (2d) show the largest plane from each equivalence class in either scan highlighted in a denser color. Rotation estimation is performed using only these highlighted planes.

where $\mathbf{UDV}^\top$ is a singular value decomposition of $\mathbf{N}_j\mathbf{N}_i^\top$, $\mathbf{D} = diag(d_i), d_i \geq d_j \; \forall \; i > j$,

$$\mathbf{S} = \begin{cases} \mathbf{I} & \text{if } det(\mathbf{N}_j\mathbf{N}_i^\top) \geq 0 \\ diag(1, 1, -1) & \text{if } det(\mathbf{N}_j\mathbf{N}_i^\top) < 0. \end{cases}$$

However, this method is unreliable if we have only two corresponding normals between scans, as shown by Eggert *et al*. [4]. For such scenarios, we follow the approach developed by Horn *et al*. [12] and later modified by Eggert *et al*. [4] to estimate the rotation between a pair of corresponded 3D points. As stated earlier, the same rotation applies between the corresponded plane normals, hence the estimate $\mathbf{R}_{ij}$ in our case is given by

$$\mathbf{R}_{ij} = \mathbf{N}_j\mathbf{N}_i^\top\mathbf{Q} \pm \frac{\mathbf{Z}}{\sqrt{|trace(\mathbf{Z})|}} \tag{13}$$

where

$$\mathbf{Q} = \left(\frac{\mathbf{u}_1\mathbf{u}_1^\top}{\sqrt{\lambda_1}} + \frac{\mathbf{u}_2\mathbf{u}_2^\top}{\sqrt{\lambda_2}}\right),$$
$$\mathbf{Z} = \begin{bmatrix}(\mathbf{N}_j\mathbf{N}_i^\top\mathbf{S})(\mathbf{N}_j\mathbf{N}_i^\top\mathbf{S})^\top & -\mathbf{I}\end{bmatrix}\mathbf{u}_3\mathbf{u}_3^\top$$

and $\{\lambda_j\}$, $\{\mathbf{u}_j\}$ are the eigenvalues and corresponding eigenvectors of the matrix $\mathbf{N}_i\mathbf{N}_j^\top\mathbf{N}_j\mathbf{N}_i^\top$. The sign in Equation (13) is chosen such that $det(\mathbf{R}_{ij}) = 1$.

4.3.2 Translation Estimation

A non-trivial least-squares solution of the optimal translation as obtained from Equation (6) requires at least three corresponding planes to exist across scans. However, real world scenes frequently have only two planes in scans, which makes the above approach ambiguous. In such cases, we use the estimated rotations to make the scans parallel to each other and employ a simplified version of the point based ICP algorithm to solve only for the underlying translation. This simplified ICP approach for pure translation converges to a locally optimal solution extremely fast.

4.3.3 Motion Estimation using Non-Planar Regions

The complete motion estimation procedure stated above works only when we have at least two non-parallel planes corresponded between scans. In practice, we have cases of non-correspondence of planes between scans or a lack of planes altogether in the scans. In such scenarios, we use a robust modification of Umeyama's point-correspondence based motion estimation method [22] to estimate the motions for all such scan pairs.

4.4. Global Registration

Once we have estimated motions between scan or equivalently their corresponding camera pairs, we can generate a viewgraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ for the given set of scans or correspondingly cameras, where for each camera pair $\{i, j\} \subseteq \mathcal{V}$, the edge $(i, j) \in \mathcal{E}$ represents that the pairwise motion $\mathbf{M}_{ij}$ has been estimated between them. Note that in practice certain edges (i, j) may be non-existent. Our objective is to obtain the absolute motion $\mathbf{M}_k$ of each vertex k of the viewgraph in some global frame of reference. It is straightforward to show that for all camera pairs $\{i, j\}$, the absolute motions $\mathbf{M}_i$ and $\mathbf{M}_j$ are related to the pairwise motion $\mathbf{M}_{ij}$ as

$$\mathbf{M}_{ij} = \mathbf{M}_j\mathbf{M}_i^{-1} \tag{14}$$

We can clearly see that to produce a minimal solution for the absolute motions from Equation (14), we need a spanning tree in $\mathcal{G}$. However, the motion estimates on the edges (pairwise motions) will, in general, be noisy, leading to unconstrained errors in the minimal solution. In order to constrain such errors, we require at least one cycle in $\mathcal{G}$. Given a minimum of N such pairwise motions for a set of N cameras, we follow a robust modification of the motion averaging approach of Govindu [8] to compute the absolute motions of each of the N cameras with respect to a global frame of reference, which is typically chosen (without loss of generality) such that the origin is fixed to the center of one of the cameras.

5. Results

5.1. Datasets

- **Augmented ICL-NUIM Dataset.** The original ICL-NUIM dataset is based on the synthetic environments provided by Handa *et al*. [10]. The availability of ground truth surface geometry enables precise measurement of trajectory estimates and reconstruction accuracy. The dataset includes around 2500-2800 clean and noisy depth images each of four models of indoor environments: two living rooms and two offices, all of dimensions roughly $5 \times 5 \times 3$ m^3. The average trajectory length of the available scenes is 36 meters for translation, while rotation covers the entire 360°. Choi *et al*. [3] developed an augmented version of this dataset that includes additional views of the scenes. We conduct our experiments on this augmented dataset.

- **Sun3D Dataset.** The Sun 3D dataset developed by Xiao *et al*. [25] consists of sequences of scans of real world indoor 3D spaces captured using standard RGB-D sensors. They provide estimates of the camera trajectories for the captured scans computed using SfM, which we use as a reasonable proxy for the ground truth for our comparison procedures. The sequences in the dataset vary in size from 1000 to 8000 scans, collected from around 10 different locations. Out of these, we have chosen 4 sequences collected from 4 different locations and consisting of scans with primarily planar regions to test our method.

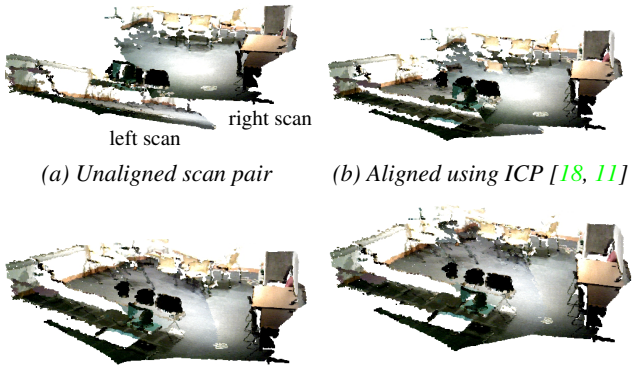
5.2. Experiments

We first present our pairwise motion — rotation followed by translation — estimate results on the datasets as per the following methodology:

- *Rotation deviation:* We compute the deviation of our estimated rotation from the ground truth rotation ($\mathbf{R}_{GT}$) in terms of the angle and the axis deviations. For a general estimate $\mathbf{R}_{est}$, angle deviation is defined as the 3D angle of the rotation matrix $\Delta\mathbf{R} = \mathbf{R}_{GT}\mathbf{R}_{est}^{-1}$, and axis deviation is simply the angle between the 3D rotation axes of $\mathbf{R}_{GT}$ and $\mathbf{R}_{est}$.

- *Translation deviation:* We compute the deviation of our estimated translation from the ground truth translation ($\mathbf{t}_{GT}$) in terms of the norm difference and the heading deviation. For a general estimate $\mathbf{t}_{est}$, norm difference is given by $\|\mathbf{t}_{GT} - \mathbf{t}_{est}\|$, and heading deviation is the angle between the translations, *i.e*., $cos^{-1}\left(\frac{\mathbf{t}_{GT}}{\|\mathbf{t}_{GT}\|} \cdot \frac{\mathbf{t}_{est}}{\|\mathbf{t}_{est}\|}\right)$.

We compare our results with results obtained using the standard point based ICP algorithm implemented in PCL [18, 11] augmented with a global motion averaging step same as that used in our method, and Zhou *et al*.'s method [27] that is currently the fastest and most accurate registration method based on point correspondences. Other global registration models such as Kintinuous [24] have already been shown to be outperformed by Choi *et al*.'s method [3], which, in turn, has been further improved upon by Zhou *et al*. [27]. Hence we do not list the performance of such methods for economy of space. Rotation estimates are given in Table 1 and translation estimates in Table 2. Note that rotation axis deviation and translation heading deviation do not affect the registration procedure as critically as the corresponding rotation angle deviation and translation norm difference if the latter are sufficiently small. We observe that our method for pairwise motion estimation has lower deviation from the ground truth compared to both ICP and Zhou *et al*. [27], which are point based methods. A primary reason for this is that the fitted plane normals are accurate since the fitting process smooths out the sensor noise in individual points. Figure 3 provides a visual representation of the registration alignment achieved by the different methods on a sample pair from the sequence `mit_32_d507/d507_2`. Our method is able to produce an alignment at par with that of Zhou *et al*. [27] and perceptibly better than ICP.



(a) Unaligned scan pair *(b) Aligned using ICP [18, 11]*

(c) Aligned using Zhou et al. *[27]* *(d) Aligned using our method*

Figure 3: Registration alignment achieved by the different methods on a sample pair of scans from the sequence `mit_32_d507/d507_2`

Next, we report the mean distances of the reconstructed surfaces to the corresponding ground truth models for the various data sequences along with the respective running times in Table 3. We calculated the mean distance to ground truth surface using the open source CloudCompare software [7] on the full reconstructions as per the instructions of Choi *et al*. [3]. The reconstruction accuracy yielded by our method matches that of Zhou *et al*. [27] and is appreciably better than that given by ICP. However, our method avoids

Table 1: Rotation deviation with respect to ground truth for pairwise motions estimated by the different methods, reported as mean angle and axis deviations in DEGREES

Dataset		ICP [18, 11]		Zhou *et al.* [27]		Our method	
		Angle deviation	Axis deviation	Angle deviation	Axis deviation	Angle deviation	Axis deviation
Augmented ICL-NUIM	livingroom1_clean	0.363	4.538	0.096	2.863	**0.094**	**2.704**
	livingroom2_clean	0.125	4.432	0.094	3.854	**0.091**	**3.666**
	office1_clean	0.243	3.526	0.118	2.311	**0.066**	**2.041**
	office2_clean	0.183	4.093	0.107	2.544	**0.075**	**2.416**
	livingroom1_noisy	0.588	7.378	0.350	4.264	**0.292**	**3.919**
	livingroom2_noisy	0.436	7.931	0.357	4.636	**0.221**	**4.305**
	office1_noisy	0.510	6.917	0.348	3.874	**0.234**	**3.791**
	office2_noisy	0.594	8.457	0.305	4.362	**0.222**	**4.127**
Sun3D	brown_bm_6/brown_bm_6	0.141	4.617	0.129	3.351	**0.106**	**2.957**
	harvard_c8/hv_c8_3	0.815	6.188	0.438	4.447	**0.414**	**4.416**
	hotel_barcelona	0.452	6.141	0.259	4.343	**0.238**	**4.176**
	mit_32_d507/d507_2	0.170	7.585	0.104	3.980	**0.086**	**3.791**

Table 2: Translation deviation with respect to ground truth for pairwise motions estimated by the different methods, reported as mean norm difference in METERS and mean heading deviation in DEGREES

Dataset		ICP [18, 11]		Zhou *et al.* [27]		Our method	
		Norm Difference	Heading deviation	Norm Difference	Heading deviation	Norm Difference	Heading deviation
Augmented ICL-NUIM	livingroom1_clean	0.018	3.534	0.008	2.268	**0.006**	**2.142**
	livingroom2_clean	0.022	3.732	0.007	2.310	**0.005**	**2.246**
	office1_clean	0.009	3.462	0.008	2.815	**0.005**	**2.721**
	office2_clean	0.009	2.842	0.005	2.473	**0.004**	**2.205**
	livingroom1_noisy	0.024	7.488	0.020	4.439	**0.015**	**4.359**
	livingroom2_noisy	0.028	7.462	0.016	4.323	**0.011**	**3.984**
	office1_noisy	0.022	6.568	0.015	3.962	**0.012**	**3.675**
	office2_noisy	0.024	6.426	0.014	4.251	**0.008**	**3.948**
Sun3D	brown_bm_6/brown_bm_6	0.036	6.536	0.032	3.742	**0.023**	**3.531**
	harvard_c8/hv_c8_3	0.019	3.991	0.017	2.957	**0.015**	**2.463**
	hotel_barcelona	0.017	6.184	0.013	3.256	**0.010**	**3.117**
	mit_32_d507/d507_2	0.024	5.747	0.016	2.962	**0.014**	**2.694**

iteratively refined nearest neighbor based as well as feature based dense point matching in majority of the scenes by exploiting the presence of planes, and is therefore significantly faster than both these approaches. Additionally, the memory usage of our method is orders of magnitude smaller than the point based methods as the input scans contain at most 8 to 10 3D planes. In contrast, the approach of Zhou *et al.* [27] uses between 1,000 to 3,000 matched points per scan pair.

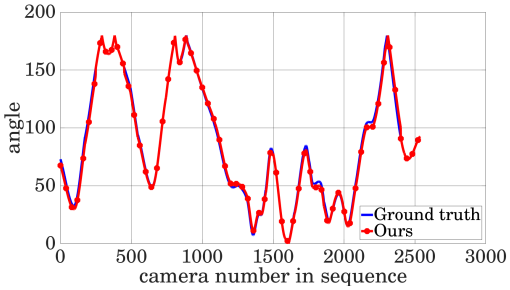
Finally, we show the camera rotation and translation trajectories recovered by our method and the corresponding ground truth trajectories for one of the sequences, namely `office2_noisy`, in Figure 4. The overall angle deviation of the recovered rotation is at most 2.5°, and the overall deviation of the recovered translation along the three coordinates is at most 0.12 m. Trajectories recovered by our method for the other sequences are similarly close to the respective ground truth trajectories.

6. Conclusion

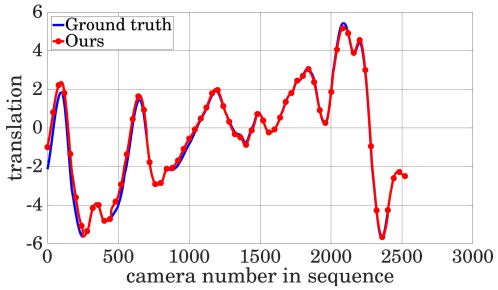
We have presented a complete registration method that exploits the presence of planar regions in 3D scans. It performs well in scenes that have an adequate number of planes. Apart from careful consideration of rotation and translation estimation between pairs of scans, we also uti-

Table 3: Mean distance from ground truth surface for full reconstruction of the individual data sequences in METERS and running times of the complete registration algorithms on the datasets, measured on an Intel Core i7-5960X 3 GHz processor with 32 GB RAM in SECONDS

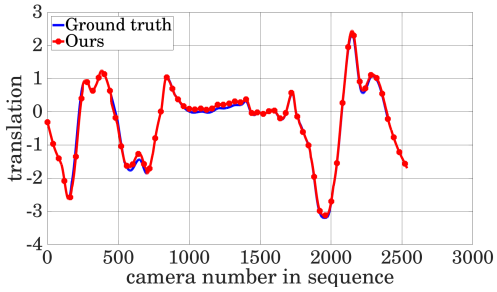
Dataset		Mean distance			Running time		
		ICP [18, 11]	Zhou *et al.* [27]	Our method	ICP [18, 11]	Zhou *et al.* [27]	Our method
Augmented ICL-NUIM	livingroom1_clean	0.06	**0.04**	**0.04**	13,800	7,175	**1,270**
	livingroom2_clean	0.04	**0.03**	**0.03**	11,300	5,875	**790**
	office1_clean	0.05	**0.02**	**0.02**	12,910	6,720	**1,100**
	office2_clean	0.05	**0.03**	**0.03**	12,220	6,350	**810**
	livingroom1_noisy	0.10	**0.05**	**0.05**	14,260	7,460	**1,380**
	livingroom2_noisy	0.07	**0.06**	**0.06**	11,680	6,110	**970**
	office1_noisy	0.08	**0.03**	**0.03**	13,370	6,990	**1,270**
	office2_noisy	0.08	**0.05**	**0.05**	12,620	6,600	**1,050**
Sun3D	brown_bm_6/brown_bm_6	0.10	**0.06**	**0.06**	5,300	2,830	**750**
	harvard_c8/hv_c8_3	0.07	**0.05**	**0.05**	4,900	2,620	**950**
	hotel_barcelona	0.08	**0.05**	**0.05**	13,280	7,110	**2,540**
	mit_32_d507/d507_2	0.07	**0.04**	**0.04**	26,590	14,240	**5,500**



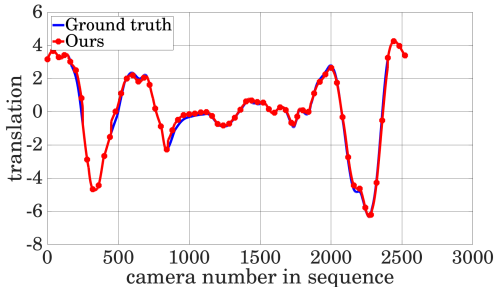
(a) Ground truth angles of rotations and our recovered angles of rotations in DEGREES



(b) Ground truth translations and our recovered translations along x-coordinate in METERS



(c) Ground truth translations and our recovered translations along y-coordinate in METERS



(d) Ground truth translations and our recovered translations along z-coordinate in METERS

Figure 4: Full camera trajectory recovered by our method plotted along side the corresponding ground truth camera trajectory

lize a robust motion averaging step that efficiently averages the relative motion estimates to provide a global solution for the 3D registration of all scans. While our method performs at par with the state-of-the-art approaches in the literature, owing to our use of 3D planes in the scans, our method is significantly faster. Moreover, since the planar representations adequately summarize the 3D information of many points, our approach results in significantly smaller memory requirements.

7. Acknowledgement

This work is supported in part by an extramural research grant by the Science and Engineering Research Board, DST, Government of India.

References

[1] H. Almohamad and S. O. Duffuaa. A Linear Programming Approach for the Weighted Graph Matching Problem. *IEEE Transactions on PAMI*, 15(5):522–525, 1993. 4

[2] A. Chatterjee and V. M. Govindu. Noise in Structured-Light Stereo Depth Cameras: Modeling and its Applications. https://arxiv.org/abs/1505.01936, 2015. 3

[3] S. Choi, Q.-Y. Zhou, and V. Koltun. Robust Reconstruction of Indoor Scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 6

[4] D. Eggert, A. Lorusso, and R. Fisher. Estimating 3-D Rigid Body Transformations: A Comparison of Four Major Algorithms. *Machine Vision and Applications*, 9(5):272–290, 1997. 5

[5] C. Feng, Y. Taguchi, and V. R. Kamat. Fast Plane Extraction in Organized Point Clouds Using Agglomerative Hierarchical Clustering. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6218–6225, May 2014. 1

[6] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann. Robust Global Registration. In *Symposium on Geometry Processing*, volume 2, page 5, 2005. 4

[7] D. Girardeau-Montaut. CloudCompare (Version 2.6.0) [GPL Software], 2014. 6

[8] V. M. Govindu. Lie-Algebraic Averaging for Globally Consistent Motion Estimation. In *CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2004. 2, 5

[9] V. M. Govindu. Robustness in Motion Averaging. In *Proceedings of the 7th Asian Conference on Computer Vision - Volume Part II*, ACCV'06, pages 457–466, Berlin, Heidelberg, 2006. Springer-Verlag. 2

[10] A. Handa, T. Whelan, J. McDonald, and A. Davison. A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM. In *IEEE ICRA*, Hong Kong, China, May 2014. 6

[11] D. Holz, A. Ichim, F. Tombari, R. Rusu, and S. Behnke. A Modular Framework for Aligning 3D Point Clouds-Registration with the Point Cloud Library. *Robotics & Automation Magazine, IEEE*, 22(4):110–124, 2015. 6, 7, 8

[12] B. K. Horn, H. M. Hilden, and S. Negahdaripour. Closed-Form Solution of Absolute Orientation Using Orthonormal Matrices. *JOSA A*, 5(7):1127–1135, 1988. 5

[13] T. K. Lee, S. Lim, S. Lee, S. An, and S. y. Oh. Indoor Mapping Using Planes Extracted from Noisy RGB-D Sensors. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1727–1733, Oct 2012. 1

[14] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-Time Dense Surface Mapping and Tracking. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '11, pages 127–136, Washington, DC, USA, 2011. IEEE Computer Society. 2

[15] K. Pathak, A. Birk, N. Vaskevicius, and J. Poppinga. Fast Registration Based on Noisy Planes with Unknown Correspondences for 3-D Mapping. *IEEE Transactions on Robotics*, 26(3):424–441, June 2010. 2

[16] K. Pathak, N. Vaskevicius, and A. Birk. Uncertainty Analysis for Optimum Plane Extraction from Noisy 3D Range-Sensor Point-Clouds. *Intelligent Service Robotics*, 3(1):37, 2009. 1

[17] S. Rusinkiewicz and M. Levoy. Efficient Variants of the ICP Algorithm. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, pages 145–152, 2001. 2

[18] R. B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011. 6, 7, 8

[19] R. Singh, J. Xu, and B. Berger. Pairwise Global Alignment of Protein Interaction Networks by Matching Neighborhood Topology. In *Research in computational molecular biology*, pages 16–31. Springer, 2007. 4

[20] Y. Taguchi, Y.-D. Jian, S. Ramalingam, and C. Feng. Point-Plane SLAM for Hand-Held 3D Sensors. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5182–5189, May 2013. 2

[21] S. Umeyama. An Eigendecomposition Approach to Weighted Graph Matching Problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):695–703, Sep 1988. 4

[22] S. Umeyama. Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Transactions on PAMI*, 13(4):376–380, Apr. 1991. 2, 3, 4, 5

[23] J. Weingarten and R. Siegwart. 3D SLAM Using Planar Segments. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3062–3067, Oct 2006. 1, 2

[24] T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. Leonard, and J. McDonald. Kintinuous: Spatially Extended KinectFusion. In *RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, Sydney, Australia, Jul 2012. 2, 6

[25] J. Xiao, A. Owens, and A. Torralba. Sun3D: A Database of Big Spaces Reconstructed Using SFM and Object Labels. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1625–1632, 2013. 6

[26] M. Zaslavskiy, F. Bach, and J.-P. Vert. A Path Following Algorithm for the Graph Matching Problem. *IEEE Transactions on PAMI*, 31(12):2227–2242, 2009. 4

[27] Q.-Y. Zhou, J. Park, and V. Koltun. Fast Global Registration. In *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*, pages 766–78, Cham, 2016. Springer International Publishing. 2, 6, 7, 8