



**HAL**  
open science

# Global Point-to-hyperplane ICP: Local and Global Pose Estimation by Fusing Color and Depth

Fernando I Ireta Munoz, Andrew I. Comport

► **To cite this version:**

Fernando I Ireta Munoz, Andrew I. Comport. Global Point-to-hyperplane ICP: Local and Global Pose Estimation by Fusing Color and Depth. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Nov 2017, Daegu, South Korea. hal-01636214

**HAL Id: hal-01636214**

**<https://hal.science/hal-01636214>**

Submitted on 2 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Global Point-to-hyperplane ICP: Local and Global Pose Estimation by Fusing Color and Depth

Fernando I. Ireta Muñoz<sup>1</sup> and Andrew I. Comport<sup>2</sup>

**Abstract**—RGB-D view registration has been widely studied by the robotics and computer vision community. The well known Iterative Closest Points (ICP) method and its variants prevail for estimating the relative pose between sensors. However, the optimization is performed locally and by consequence it can get trapped in local minima. Global registration methods have been introduced as an approach to solve the local minima problem by exploiting the geometric structure of SE(3), and accelerated with local approaches. In this paper, a local hybrid approach named Point-to-hyperplane ICP has been combined with a global Branch and Bound strategy in order to estimate the 6DOF (degrees of freedom) pose parameters. Registration is performed by considering color and geometry at both, the matching and the error minimization stages. Results in real and synthetic environments demonstrate that the proposed method can improve global registration under challenging conditions such as partial overlapping and noisy datasets.

## I. INTRODUCTION

A fundamental problem in RGB-D registration is estimating the pose that relates two sets of measurements obtained by a moving sensor (or sensors) at different times. This view registration problem has been investigated extensively by the robotics and computer vision community and employed for many applications such as 3D reconstruction, autonomous navigation and visual SLAM.

When RGB-D sensors are employed, color and depth can be registered simultaneously at a considerably high frame-rate, by local registration methods that improve the alignment between extended frames (RGB + Depth). Local techniques, however, consider a small relative motion between the frames and a large overlap, leading to local minima when the registered frames are not close enough. When a small overlap is present between datasets, registration can be performed using global approaches. Global registration approaches can obtain rough alignments, regardless of the initial pose between the frames. However, since they are computationally expensive, global methods are often only used to initialize local approaches. The combination of global and local approaches guarantees however, the convergence of the non-convex error function thereby dealing with the issue of local minima.

Local RGB-D pose estimation from color and depth images have each been individually studied for image-based and geometric-based approaches, respectively. In the case of *geometric* approaches, the well known Point-to-point Iterative Closest Point (ICP) [1] algorithm and its variants [24]

have demonstrated a good performance for aligning 3D pointclouds. In particular, the Point-to-plane and the Plane-to-plane ICP strategies [4], [25] have been shown to improve the robustness and convergence of the classic ICP algorithm. On the other hand, *image-based* approaches can estimate the pose by iteratively minimizing the photometric error [5]. Various prominent feature-based methods first extract geometric information from the image before performing registration over the geometric error. This requires, however, error prone extraction of features. An alternative is found in direct methods, which perform registration based on image warping [5]. Finally, *hybrid* methods have demonstrated more robust and accurate performance than *image-based* or *geometric-based* methods alone. *Hybrid* approaches can obtain the main advantages of image-based and geometric-based approaches during the minimization process by minimizing both errors simultaneously. Various hybrid methods are presented and cited in [29], [15], [14]. Hybrid methods differ in how the closest points are estimated and how the uncertainty factor between color and depth is determined.

Global registration approach have been introduced to obtain a coarse alignment between scan pairs that are not roughly aligned. One of the simplest global approaches consist in shifting the centroids of the pointclouds to the origin of the coordinate system and statistical methods such as the Principal Component Analysis (PCA) can estimate an initial orientation by extracting strong patterns in a dataset. The PCA algorithm has limitations for 3D registration, such as a 180 degrees ambiguity or eigenvalues similar in value (axes may switch). In the literature, global 3D pre-alignment is usually performed by matching geometric features from the scene. Extracted features can be useful to obtain a faster initial alignment between the datasets, but it can require a manual or an automatic selection of good correspondences and a robust estimator to reject outliers (e.g. RANSAC).

A global registration method that pre-selects correspondences by using the Fast Point Feature Histogram (FPFH) is presented in [30]. This method obtains a fast alignment by aligning only good pre-computed correspondences in the iterative loop without using a local approach for refinement. Approaches that use both, global and local registration with extracted features, can be found in [12], [7], [18], and application to RGB-D registration can be found in [8]. [12] performs registration by finding a so-called model graph, which contains a connection between correct matches. A global approach named Super 4PCS [20] uses a sub-sampling type strategy to extract coplanar 4-points sets, which are assumed to remain invariant under rigid motion. The 4-

\*This work is supported by the European H2020 project: COMANOID, Université Côte d'Azur, CNRS, I3S, France and CONACYT, México.

<sup>1</sup> ireta@i3s.unice.fr <sup>2</sup> Andrew.Comport@cnrs.fr

point sets are created in all scan pairs and the algorithm finds all equivalent 4-points in between. An approach based on the Branch-and-bound (BnB) method is proposed in [7]. The method is improved by finding potential corresponding points based on performing integral operations on the underlying shape. In [18], the transformation is globally estimated by maximizing the correlation between extended Gaussian images (EGIs) in the Fourier domain. The method uses the convolution of the spherical harmonics of the EGIs and the rotational Fourier transform as features to estimate rotation. The translation is found similarly by employing fast Fourier transform. A RGB-D application is found in [8], the global estimation is found by minimizing the correspondences of extracted linear edges along creases and contours in depth images.

Methods that do not require any feature detection, such as the Branch and Bound algorithm [17], can be employed to estimate the pose by exploring and sub-dividing all rotational and translational space as in [10], [2], [12], [22], [28], [3], [26]. These methods, however, are vulnerable to outliers and partial overlap. Various methods like [10], [2], [22] consider pure rotation for the BnB method since the translation is known apriori or the clouds have sufficient overlap to be estimated by the local approach. In [22], the rotation space is sub-divided in its quaternion representation and the translation is estimated by variants of local approaches [11]. An optimized version of [10] has been presented in [2]. Both strategies use the axis-angle representation to parametrize a solid radius- $\pi$  sphere for rotations. The BnB method sub-divides the sphere by using an octree data structure, where the optimal global solution is found by bounding the generated subsets. Strategies such as [28], [3], [26] have extended the exploration of the transformation space by subdividing the rotational and translational space together. These methods differ in:

- how the  $SE(3)$  space is sub-divided (branching).
- how the upper and lower bounds are computed.

As in [10], the Go-ICP [28] and GOGMA [3] algorithms propose an axis-angle representation to sub-divide the  $SO(3)$  space, where an inner BnB search structure is included to explore the translational space. Whilst Go-ICP minimizes the  $L_2$  norm of the Point-to-point residual error, GOGMA uses a Gaussian mixture alignment (GMA), which is less sensitive to partial overlap. Both methods estimate the uncertainty radius for rotation and translation, which is the key for estimating the lower bounds of the error function. The computational cost is linear with the size of datasets and it is improved by a subsampling stage. Recently, a more efficient method has been proposed in [26]. The method follows the same pipeline of minimizing the  $L_2$  norm of the residual errors, but the BnB exploration decouples the rotational and translational space via the use of surface normals. The distribution of the surface normals are modeled as a von-Mises-Fisher mixture model (vMF-MM) for rotation and as a Gaussian Mixture Model for translation.

However, to the best of our knowledge, global and local

approaches in the literature have been estimating the pose by considering only 2D and 3D geometry. In this paper, a method that also integrates color as an available criteria in both, matching and error minimization stages, is proposed. The method combines a variant of the BnB method presented in [28] in higher dimensions with hybrid approaches [15], [19] as the local refinement. The proposed method has been applied in a visual SLAM and dense mapping setting, where mistracking is a common failure mode due to incorrect estimations of the pose or small overlap between acquired frames. Results in both, real and synthetic environments, will demonstrate that 3D global methods can be extended to RGB-D global pose estimation and accelerate convergence.

The article is structured as follows: In Section II, a brief introduction is given about how global and local approaches can estimate the pose and how both strategies can improve registration. Particularly, iterative registration between the BnB algorithm and local ICP approaches will be explained in this Section. The proposed Point-to-hyperplane ICP approach will be presented in Section III. In this Section it will be demonstrated how the global registration can be improved when color is considered as an available criterion for the minimization process by using hybrid approaches as the local refinement. Finally, results in both, real and simulated environment, will be shown in Section IV.

## II. GLOBAL AND LOCAL APPROACHES FOR POSE ESTIMATION

Global pose estimation approaches can obtain a pre-alignment without making assumptions about the relative initial pose or the overlap ratio between datasets. Global approaches are often employed to initialize local approaches, which refine the convergence. Recently, the BnB method has been used in collaboration with the ICP algorithm to improve 3D registration. Whilst the BnB algorithm provides a globally optimal solution by exploring the entire transformation space (all feasible solutions), the ICP algorithm refines the local alignment (optimal solution). The proposed approach in this paper is based on the Go-ICP algorithm, which only considers 3D registration. However, a generalized framework of the Global BnB method and local ICP approaches will be introduced here and subsequently extended to higher dimensions. For the purposes of this paper, global registration strategies based on BnB methods, which subdivide the  $SE(3)$  transformation space, will be considered.

The minimization of the sum of least squared error  $L_2$  objective function will be considered:

$$\mathbf{E}(\mathbf{T}(\mathbf{x})) = \sum_{i=1}^N \mathbf{e}_i^2 \quad (1)$$

where each  $\mathbf{e}_i$  error metric refers to the  $N$  correspondences between two measurement sets, which are acquired at different views (they can be transformed via the function  $f(\cdot)$ ), as:

$$\mathbf{e}_i = \mathbf{M}_i^* - f(\mathbf{M}_i, \mathbf{x}) \in \mathbb{R}^{nD} \quad \forall i \in N \quad (2)$$

A  $nD$ -vector  $\mathbf{M}_i \in \mathbb{R}^n$  ( $n$ -dimensional) is defined here as a vector that contain all the measurements obtained by the

sensor (or sensors) at the  $i$ -th point. The superscript  $*$  will be used throughout this paper to identify the measurements that were obtained first (reference dataset). The homogeneous pose matrix  $\mathbf{T}(\mathbf{x}) \in \mathbb{R}^{4 \times 4} = (\mathbf{R}(\mathbf{x}), \mathbf{t}(\mathbf{x})) \in SE(3)$  can be decomposed in the rotational  $\mathbf{R}(\mathbf{x}) \in SO(3)$  and translational  $\mathbf{t}(\mathbf{x}) \in \mathbb{R}^3$  components, which depend on a minimal parametrization of the linear and angular velocity  $\mathbf{x} = [v, \omega]^T$ , respectively. Both components are related via the exponential map as  $\mathbf{T}(\mathbf{x}) = e^{[\mathbf{x}]_{\wedge}}$ , with the operator  $[\cdot]_{\wedge}$  as:

$$[\mathbf{x}]_{\wedge} = \begin{bmatrix} [\omega]_{\times} & v \\ \mathbf{0} & 0 \end{bmatrix} \quad (3)$$

where  $[\cdot]_{\times}$  is the skew symmetric matrix operator.

The 6 DOF pose parameter  $\mathbf{x}$  can be estimated by minimizing the error function (2). Two cases can be found while estimating the pose: If the rotation and translation are known (case 1), then the correspondences can be estimated or if correspondences are known (case 2), then the rotation and translation can be estimated. In any case, (2) becomes solvable. However, erroneous alignments can be obtained when wrong correspondences are found (vulnerability to outliers). Local minima are problematic for local estimation approaches. Despite this, local approaches are usually faster than global approaches and they can obtain robust alignments if the datasets are close enough. On the other hand, when a BnB method is used, only a rough alignment can be obtained between frames due to the fact that an extensive exploration of the  $SE(3)$  transformation space is prohibitively inefficient and memory consuming. The aim of fusing global and local methods for registration can be seen at this point: The non-convex function (1) can be optimally solved if the BnB method and the ICP algorithm work together. In order to guarantee convergence and avoid local minima, each method can be initialized with the result of the other until convergence (See Fig. 1).

In order to introduce how the global BnB and local ICP approaches find the transformation  $\mathbf{T}(\mathbf{x})$  that minimizes (1), a 6DOF pose estimation framework for each will be introduced separately.

#### A. Local registration framework

*Geometric-based* or *photometric-based* pose estimation have been widely studied. Since they share much similarity in estimating the unknown pose parameter  $\mathbf{x} \in \mathbb{R}^6$  in (2), they can be performed using a non-linear iteratively re-weighted least squares (IRLS) method. Generally, the IRLS method encompasses the following stages:

- 1) Acquire a set of measurements (e.g. color and depth in case of RGB-D sensors) at two different viewpoints.
- 2) Transform the current measurements onto the reference measurements using the current pose estimate.
- 3) Find correspondences between the transformed and the reference measurement vectors.
- 4) Minimize the weighted error function and estimate a pose update and repeat to 2) until convergence.

Given a particular sensor modality, methods in the literature essentially differ in how the closest points are estimated

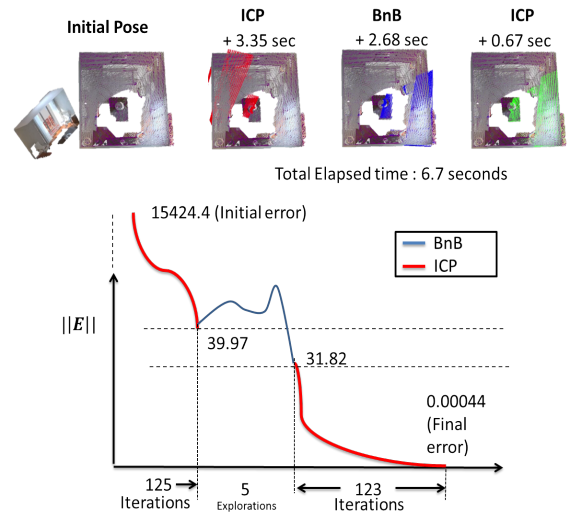


Fig. 1. Iterations and elapsed time of Go-ICP 3D registration (Read Section II-C for implementation details). A subset of the cloud was extracted and transformed with a random pose. It can be seen on the top that a local minima is reached in the first trial of ICP. The BnB bypasses the local minima and obtains a closer solution, which is again refined by the ICP algorithm. The cost function evolution is represented on the bottom.

and how the error function is minimized (step 3 and 4, respectively). For the purposes of this paper, the pose update is estimated by using a robust ICP algorithm (as in classic Go-ICP [28]). The given non-linear error in (1) is minimized by using the Gauss-Newton approach, which iteratively computes the unknown pose parameter as:

$$\mathbf{x} = -(\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W} \mathbf{e} \in \mathbb{R}^6 \quad (4)$$

where  $\mathbf{J}$  is the Jacobian matrix obtained by taking the derivative of the error function (2), and the weight matrix  $\mathbf{W}$  contains the weights  $\rho_i$  obtained by a M-estimation [13]. The pose  $\mathbf{T}(\mathbf{x})$  is updated at each iteration as  $\hat{\mathbf{T}} \leftarrow \hat{\mathbf{T}} \mathbf{T}(\mathbf{x})$  until convergence.

#### B. Global registration framework

The BnB algorithm is used to solve non-convex problems by searching the complete space of solutions for the best solution. Variants of the BnB method have been employed for 3D registration. In order to provide a generalized framework for pose estimation using only the BnB algorithm, the following stages can be established:

- 1) Subdivide the transformation space into subspaces.
- 2) Transform the current measurement vector and estimate correspondences w.r.t. the reference measurement vector by using the centroid coordinates of each subspace.
- 3) Evaluate the upper-bounds and lower-bounds for all live subspaces.
- 4) Discard subspaces from the live subspaces by a search strategy, which selects the next subspace to process.
- 5) Repeat to 2) if a branching is performed before bounding the error, or to 1) if a branching is performed after bounding the error, until convergence.

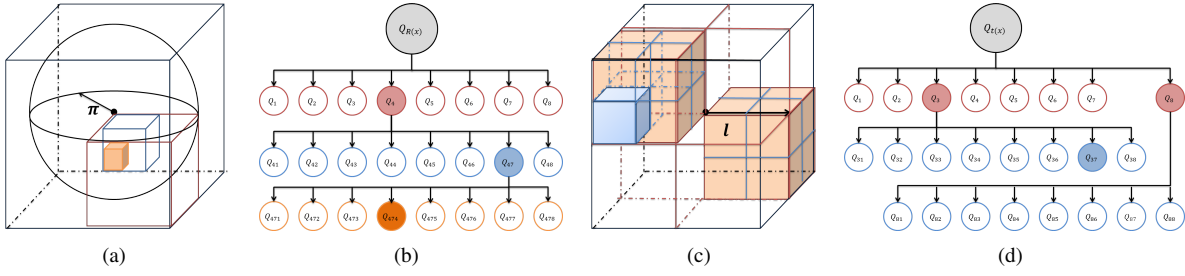


Fig. 2. Parametrization of the  $SE(3)$  space and subcubes. The rotation space can be represented in a  $\pi$ -sphere within a cube of  $2\pi$  length (See text for more detailed information) (a) and a 3D cube half side-length  $l$  for translation (c). An example of an octree branching is presented in (b) and (d). The subspace candidates that may contain the global solution are colored, its correspondent subspace is represented in (a) and (d).

Global registration methods can differ in how branching, bounding and subspace searching are performed. In the following, all three components will be briefly introduced.

**Branching the  $SE(3)$  space:** All branching rules in the context of BnB can be seen as a subdivision of the search space through the addition of constraints. The subdivision method can be performed before or after bounding the error function (the strategies are so-called eager and lazy BnB, respectively). The difference is that eager BnB subdivides the transformation space as soon as new subspaces are available (branching is performed first), whilst lazy BnB subdivides the subspace if necessary (bounds are evaluated first). In this paper, a lazy-BnB pipeline with a nested 6D branching strategy is employed. A nested BnB strategy performs an inner translation BnB after evaluating live subspaces for the rotational space. This is less memory consuming than estimating a direct 6D space BnB.

Similarly to [28] and [3], an octree structure is employed in this paper to subdivide the rotation and translation spaces. Rotation space is parametrized using the axis-angle representation as a vector  $\mathbf{r} = \alpha \hat{r} \in \mathbb{R}^3$  where  $\hat{r}$  is the axis and  $\alpha$  is the angle. Therefore, the rotational  $SO(3)$  space can be represented in a solid sphere of radius  $\pi$  [10]. The translation space can be represented as a constrained cube  $C_t = [-l, l]^3$ , within which it is assumed that the optimal translation will be found. Similarly, the  $\pi$ -sphere can be circumscribed in a 3D cube  $C_r = [-\pi, \pi]^3$ . Each generated sub-cube  $C_{R_j} \subset C_R$  and  $C_{t_j} \subset C_t$  represents a search state with possible solutions to (2) (See Fig. 2).

**Bounding the error function:** The bounding function is the fundamental component of the BnB. The upper and lower bounds of (1) are estimated for each generated subspace within  $C_R$  and  $C_t$ , and they are determined from the residual (2). For this paper, the optimal registration error was computed as:

$$\bar{e}_i = \mathbf{M}_i^* - f(\mathbf{T}(\mathbf{x}), \mathbf{M}_i) \quad (5)$$

$$e_i = \max((\mathbf{M}_i^* - f(\mathbf{T}(\mathbf{x}), \mathbf{M}_i)) - (\psi_{R_i} + \psi_{t_i}), 0) \quad (6)$$

where  $\bar{e}_i$  and  $e_i$  are the upper and lower bounds, respectively.

The lower bound is derived by considering a mathematical concept called *uncertainty radius*  $\psi = \psi_{R_i} + \psi_{t_i}$  in terms of the angle metric [10], [28]. It examines the uncertainty region of a point  $\mathbf{M}_i$  perturbed by an arbitrary rotation  $\mathbf{R}_j(\mathbf{x}) \in C_R$  or a translation  $\mathbf{t}_j(\mathbf{x}) \in C_t$  corresponding to the center

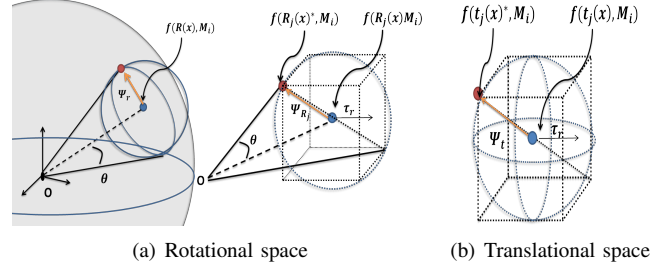


Fig. 3. Uncertainty radius at a point centered in  $\mathbf{T}(\mathbf{x})$ . a) Rotation uncertainty radius  $\psi_r$  and b) Translation uncertainty radius  $\psi_t$  are estimated by the half-side length subcube circumscribed in a sphere.

of the cubes with half-side length  $\tau_r$  and  $\tau_t$ , respectively. These uncertainty regions  $C_R$  and  $C_t$  are enclosed in a  $n$ -dimensional sphere centered at  $f(\mathbf{T}_j(\mathbf{x}), \mathbf{M}_i)$  and estimated separately for rotation and translation as is shown in Fig. 3, where  $\mathbf{T}(\mathbf{x})^* = (\mathbf{R}(\mathbf{x})^*, \mathbf{t}(\mathbf{x})^*)$  runs over all rotations and translations, and represented by points in the cubes. The maximum length of the uncertainty radius is obtained by considering the maximum distance between two transformations and it can be extended to higher dimensions with the following lemma, which summarizes the uncertainty radius in  $n$ -dimensions:

**Lemma 2.1:** Given a  $nD$ -point  $\mathbf{M}_i$ , a half side length  $\tau_r$  rotation cube  $C_r$  and a half side length  $\tau_t$  translation cube  $C_t$ , centered at  $\mathbf{T}(\mathbf{x})$ , then  $\forall \mathbf{R}_j(\mathbf{x}) \in C_R$  and  $\forall \mathbf{t}_j(\mathbf{x}) \in C_t$ ,

$$\|f(\mathbf{t}_j(\mathbf{x})^*, \mathbf{M}_i) - f(\mathbf{t}_j(\mathbf{x}), \mathbf{M}_i)\| \leq \sqrt{nD} \tau_t \doteq \psi_t \quad (7)$$

$$\|f(\mathbf{R}_j(\mathbf{x})^*, \mathbf{M}_i) - f(\mathbf{R}_j(\mathbf{x}), \mathbf{M}_i)\| \leq \min(\sqrt{nD} \tau_r, \pi) \doteq \psi_{R_i} \quad (8)$$

An important point from (8), is that the angle distance is less than the Euclidean distance in the angle-axis representation. Therefore, the maximum distance between two rotations cannot be bigger than the maximum angle  $\theta \leq \sqrt{3} \tau_r$  [10]. In the 3D case, the respective radius  $\psi_{R_i}$  and  $\psi_{t_i}$  are integrated into the error function to estimate the lower bound. The uncertainty radius for rotation can be estimated as follows:

$$\psi_{R_j} = 2 \sin \left( \frac{\|\mathbf{R}_j^* \mathbf{M}_i - \mathbf{R}_j \mathbf{M}_i\|}{2} \right) \|\mathbf{M}_i\| = 2 \sin \left( \frac{\sqrt{3} \tau_r}{2} \right) \|\mathbf{M}_i\| \quad (9)$$

By observing that the translation cube can be inscribed in a sphere, the uncertainty radius for translation is simply estimated as:

$$\psi_t = \|(\mathbf{M}_i + \mathbf{t}_j(\mathbf{x})^*) - (\mathbf{M}_i + \mathbf{t}_j(\mathbf{x}))\| = \sqrt{3} \tau_t \quad (10)$$

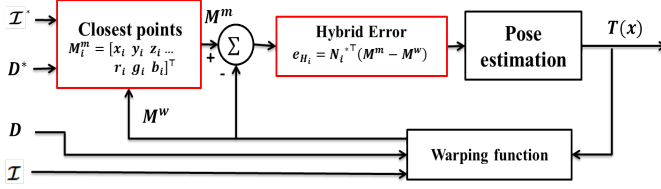


Fig. 4. Point-to-hyperplane ICP registration. The method can perform the RGB-D registration without estimating a tuning parameter between color and depth [15] and it considers the color as a criterion to find the correspondences.

**Searching a global solution:** There are three main strategies for exploring generated subspaces and selecting those which may contain an optimal solution in the BnB algorithm: *Best-first*, *breadth-first* and *depth-first* strategies. These strategies mainly differ in the priority provided to a subspace to be evaluated. In this paper, a best-first search is employed. The best-first strategy always selects the subspaces with the lowest bound, where no extra bound calculations take place after getting the optimal solution. The BnB searching stops when the difference between the globally minimal error  $e_i$  and the lower bound  $\underline{e}_i$  of the current subspace is less than an established threshold ( $e_i - \underline{e}_i < \varepsilon$ ) or when the remaining subspaces are small enough (upper bound is smaller than the globally minimal error:  $\bar{e}_i > e_i$ ).

### C. Integrating local registration in global registration

Combining BnB with ICP for registration allows to avoid local minima along with more accurate and robust alignments. In the classic BnB algorithm, an ICP method is used to refine alignment if the upper bound is smaller than the current best error ( $\bar{e}_i < e_i$ ). In this paper, the Point-to-hyperplane ICP is used while evaluating the rotation bounds in the nested BnB search. An example of 3D registration is shown in Fig 1. The BnB updates the upper bounds with the current best error  $e_i$  obtained by ICP. By performing this, more subspaces can be discarded for the BnB search. The BnB provides a subspace with a better solution which is refined by the local approach.

## III. POINT-TO-HYPERPLANE GLOBAL ICP FOR RGB-D POSE ESTIMATION

When RGB-D registration is performed, hybrid methods can be employed to simultaneously minimize the error generated between measurements. The proposed method in this paper follows the pipeline shown in Section II, but the point-to-hyperplane ICP method is considered to refine the pose estimates obtained by the BnB strategy presented in Section II-B. The proposed hybrid method (Fig. 4) has the advantage that it is invariant to the uncertainty between color and depth [14] and it can obtain robust and faster convergence than classic approaches while maintaining the benefits of the geometric and photometric methods. According to the following lemma, the tuning parameter is omitted since the method is invariant to it.

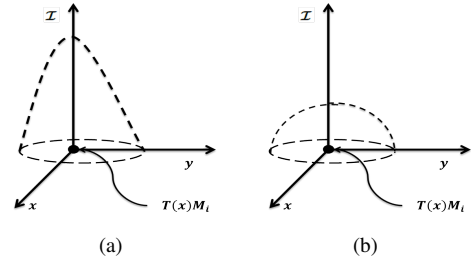


Fig. 5. Uncertainty region of 3D hybrid points  $\mathbf{M}_i = [x_i \ y_i \ I_i]^\top$ . The different metric between color and geometric points, which are perturbed by a transformation  $\mathbf{T}(\mathbf{x})$ , creates an uncertainty n-paraboloid (a). It is assumed that an uncertainty radius cannot be properly estimated from this. Therefore, the geometric and photometric measurements can be normalized to maintain a unitary half-sphere in n-dimensions (b).

*Lemma 3.1:* The integrated error  $\mathbf{e}_H$  in  $j$ -th dimension is invariant to the relative scale  $\lambda$  if it is minimized by a Point-to-hyperplane method.

$$\mathbf{e}_{H_i} = \mathbf{N}_i^{*\top} (\mathbf{M}_i^* - f(\mathbf{M}_i, \mathbf{x})) = \lambda \mathbf{N}_i^{*\top} (\mathbf{M}_i^* - f(\mathbf{M}_i, \mathbf{x}))$$

Therefore, the error function shown in (2) can be replaced with the following error function:

$$\mathbf{e}_{H_i} = \mathbf{N}_i^{*\top} (\mathbf{M}_i^m - \mathbf{M}_i^w) \in \mathbb{R}^n \quad (11)$$

where  $\mathbf{M}_i^m$  denotes corresponding matches found between the reference and the transformed current measurements.  $\mathbf{M}_i^w$  is the measurement vector transformed by the warping function  $w(\cdot)$ , which projects a reference measurement vector onto the current reference frame. For the experiments in this paper, each 3D point will be associated with a unique intensity value  $I_i$  as:  $\mathbf{M}_i = [x_i \ y_i \ z_i \ I_i]^\top$ .  $\mathbf{N}_i^* \in \mathbb{R}^4$  are the 4D normals of the reference measurements.

Upper and lower bounds cannot be easily extended to higher dimensions by analyzing (11). The challenge lies in how the uncertainty radius of a n-sphere is obtained for lower bounds. Shown in (7) and (8), the uncertainty regions in n-dimensions centered at  $f(\mathbf{T}_j(\mathbf{x}), \mathbf{M}_i)$  can be computed. However, the different scale between 3D Euclidean points and intensities changes the n-sphere into a n-paraboloid. To analyze the uncertainty radius for color and depth, a 3D scenario will be considered instead of the full 4D vector. Consider a 3D hybrid measurement vector as:  $\mathbf{M}_i = \lambda [x_i \ y_i \ I_i]^\top$ , where  $\lambda = \text{diag}(\lambda_x, \lambda_y, \lambda_I)$  is the uncertainty factor. The equation of a 3D sphere is given by  $x^2 + y^2 + z^2 = d^2$  which can be represented as a particular case of an ellipsoid as:  $\frac{x^2}{A^2} + \frac{y^2}{B^2} + \frac{z^2}{C^2} = 1$ , where the length of the semi-axes  $(A, B, C)$  is equal. For the hybrid 3-vector,  $\frac{M_x^2}{A^2} + \frac{M_y^2}{B^2} + \frac{M_I^2}{C^2} = 1$ ,  $A^2 = B^2 \neq C^2$ . Furthermore, the color axis is constrained by the minimum and maximum intensities (0-255), which constrain the ellipsoid to a paraboloid centered at the intensity value  $I_i$  as is shown in Fig. 5. In the 4D case, the hyper-ellipsoid can be written as:  $\frac{M_x^2}{A^2} + \frac{M_y^2}{B^2} + \frac{M_z^2}{C^2} + \frac{M_I^2}{D^2} = 1$ , where  $A, B, C$  and  $D$  are affected by  $\lambda = (\lambda_G, \lambda_I)$ . In order to use a similar bounding by considering the uncertainty radius, the data are normalized. The normalization of the

TABLE I

AVERAGES IN TIME (MILLISECONDS), NUMBER OF ITERATIONS FOR CONVERGENCE, RELATIVE POSE ERROR (RPE) AND ABSOLUTE TRAJECTORY ERROR (ATE). THE AVERAGE TIME SHOWN DOES NOT CONSIDER THE COMPUTATION OF THE NORMALS OR MATCHING STAGE. THE NUMBER OF TIMES THAT THE METHOD GOT TRAPPED IN LOCAL MINIMA IS SHOWN AS A RATIO  $\#localminima(Lmin)/\#keyframes(KF)$ . A RGB-D KEYFRAME WAS SELECTED AT EVERY  $n - th$  FRAME IN THE SEQUENCE WHICH IS SHOWN BELOW THE SEQUENCE NAME ALONG WITH THE NUMBER OF TOTAL OF RGB-D FRAMES IN THE SEQUENCE AS  $n - frames/total - frames$ .

Sequence	Method	ATE (m)		AVERAGE		$\frac{\#Lmin}{\#KF}$
		RMSE	MEAN	Time(sec)	#Iterations	
(10/1508)	G-P2PI	0.125	0.111	<b>0.376</b>	<b>28.20</b>	6/150
	I-ESM	0.145	0.119	0.455	30.50	12/150
	H-NA- $\lambda$	<b>0.028</b>	<b>0.026</b>	0.566	44.24	7/150
	H-A- $\lambda$ -1	0.046	0.042	1.017	77.51	<b>1/150</b>
	H-A- $\lambda$ -2	0.086	0.077	2.428	61.66	<b>1/150</b>
	GGICP	<b>0.028</b>	<b>0.026</b>	0.486	42.32	<b>1/150</b>
(10/965)	G-P2PI	0.110	0.090	<b>0.265</b>	<b>19.40</b>	12/96
	I-ESM	0.119	0.096	0.346	20.63	16/96
	H-NA- $\lambda$	<b>0.035</b>	<b>0.031</b>	0.446	32.24	11/96
	H-A- $\lambda$ -1	0.062	0.053	0.773	55.07	9/96
	H-A- $\lambda$ -2	0.113	0.081	2.165	54.96	6/96
	GGICP	<b>0.035</b>	<b>0.031</b>	0.368	30.59	<b>4/96</b>
(10/880)	G-P2PI	0.040	0.038	0.409	32.03	6/88
	I-ESM	0.040	0.038	0.490	33.05	9/88
	H-NA- $\lambda$	0.028	0.026	0.417	30.65	0/88
	H-A- $\lambda$ -1	0.527	0.492	0.938	69.81	1/88
	H-A- $\lambda$ -2	0.787	0.748	2.112	54.16	4/88
	GGICP	<b>0.018</b>	<b>0.017</b>	<b>0.337</b>	<b>29.38</b>	<b>0/88</b>
(10/1240)	G-P2PI	0.771	0.729	<b>0.376</b>	<b>28.20</b>	8/124
	I-ESM	0.815	0.777	0.455	30.50	13/124
	H-NA- $\lambda$	0.264	0.229	0.688	47.09	6/124
	H-A- $\lambda$ -1	0.368	0.353	1.191	90.47	4/124
	H-A- $\lambda$ -2	0.763	0.721	3.068	77.03	4/124
	GGICP	<b>0.231</b>	<b>0.218</b>	0.642	46.27	<b>2/124</b>

measurements allows to keep a half n-sphere volume (n-paraboloid with same length in all axes), instead of a n-ellipsoid. On the other hand, it is well known [23] that a n-sphere decrease its volume after the 5-th dimension. Therefore, it is considered that an inscribed n-cube in the n-sphere may not properly represent the uncertainty radius. For the experiments of this paper, a normalized 4-vector  $\mathbf{M}$  was considered. Therefore, the uncertainty radius of a 4-dimensional space with a maximum aperture angle  $\theta = \sqrt{N}\tau$  ( $N = 4$ ) can be defined here as:

$$\begin{aligned} \psi_{R_i} &= 2\sin(\tau_r) \|\mathbf{M}_i\| \\ \psi_l &= 2\tau_r \end{aligned} \quad (12)$$

which are employed in (6) to estimate the 4D lower bounds. The minimization of the upper and lower bounds in 4 dimensions can be chosen as:

$$\overline{\mathbf{E}(\mathbf{T}(\mathbf{x}))} = \sum_{i=1}^N \max(\mathbf{e}_{H_i}, 0)^2 \quad (13)$$

$$\underline{\mathbf{E}(\mathbf{T}(\mathbf{x}))} = \sum_{i=1}^N \max(\mathbf{e}_{H_i} - (\psi_{R_i} + \psi_l), 0)^2 \quad (14)$$

where  $\mathbf{e}_{H_i}$  is the error function in (11).

For purposes of this paper, the 4D BnB method was used to initialize the Point-to-hyperplane ICP approach when a local minima is found. The fusion of both methods can improve the convergence rate for global registration while obtaining robust alignments. The results of the proposed Global Point-to-hyperplane ICP (GP2HPI) method are shown in the following Section.

#### IV. RESULTS

In order to evaluate the GP2HPI algorithm, three experiments were performed: 1) Registration of pairs of RGB-D frames with a large initial transformation between them. 2) Keyframe visual odometry and 3) RGB-D global registration of keyframes with small overlap. A multi-resolution pyramid was employed to improve computational efficiency (resolution  $160 \times 120$  at the top). Rejection of outliers was handled with M-estimation [13]. For the matching stage, two strategies were tested: a 6-dimensional kd-tree [21] for local approaches and the Distance Transform [6] strategy for BnB. The method was compared among geometric-based, photometric-based and hybrid error functions which differ in how the tuning parameter  $\boldsymbol{\lambda} = (\lambda_G, \lambda_I)$  is estimated. Consider the following error function:

$$\mathbf{e}_i = \begin{pmatrix} \lambda_G (\mathbf{N}_i^{*T} (\mathbf{P}_i^m - \mathbf{P}_i^w)) \\ \lambda_I (I_i^m - I_i^w) \end{pmatrix} \in \mathbb{R}^4 \quad (15)$$

where  $\mathbf{P}_i^w \in \mathbb{R}^3$  is the warped 3D point and  $I_i^w$  is the warped intensity.  $\mathbf{P}_i^m \in \mathbb{R}^3$  and  $I_i^m$  are the correspondences. The variants of the hybrid methods will be identified as follows:

- 1) *G-P2PI*. Geometric Point-to-plane ( $\lambda_I = 0$ ) [4].
- 2) *I-ESM*. Direct-method ( $\lambda_G = 0$ ) [5].
- 3) *H-NA- $\lambda$* . 1) + 2) + non adaptive lambda [19].
- 4) *H-A- $\lambda$ -1*. 1) + 2) + adaptive lambda [27].
- 5) *H-A- $\lambda$ -2*. 1) + 2) + adaptive lambda [16].

For purposes of this paper, the photometric term is minimized by the Efficient Second Order minimization (ESM).

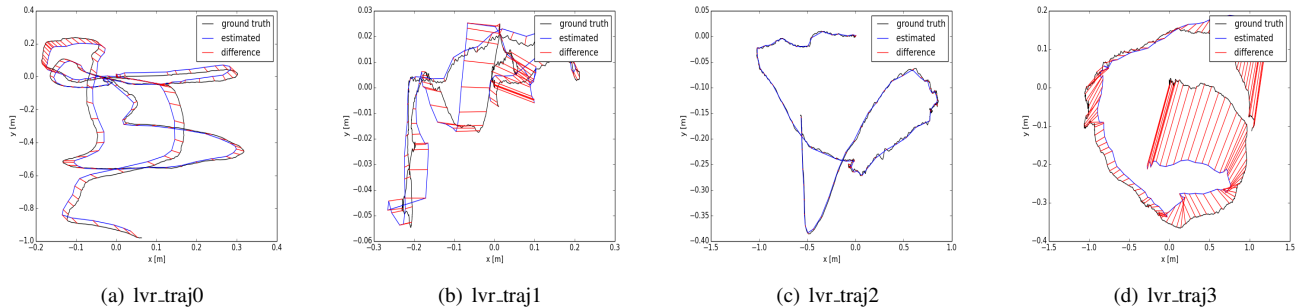


Fig. 6. Examples of the Absolute Trajectory Error evaluation obtained by the Point-to-Hyperplane method combined with the BnB method. The benchmark datasets [9] were used and a keyframe was chosen at each 10th frame of the sequence. It can be observed that closer estimations w.r.t. the groundtruth can be obtained with a refinement with the Point-to-hyperplane as the local approach while avoiding a mistracking problem.

*Experiment 1:* A synthetic RGB-D frame was employed as a reference and transformed. 100 Images were synthesized by warping the reference image with a large random pose. Initially, local methods could not obtain convergence or a local minima was reached, but they could align the frames when they were initialized with the solution obtained by the BnB algorithm. The performance of local methods after being initialized by the BnB is presented in Table II.

TABLE II  
AVERAGES IN TIME AND IN NUMBER OF ITERATIONS UNTIL CONVERGENCE FOR 100 SYNTHESIZED IMAGES AT RANDOM POSES. THE RGB-D IMAGE WAS GENERATED FROM A SYNTHETIC ENVIRONMENT, WHERE GAUSSIAN NOISE WAS ADDED. THE LOCAL APPROACHES WERE INITIALIZED WITH THE BnB METHOD.

Method	# Iterations	Time (sec)
1) <i>G-P2PI</i> [4]	154.34	1.8144
2) <i>I-ESM</i> [5]	144.12	1.6731
3) <i>H-NA-<math>\lambda</math></i> [19]	115.77	1.3472
4) <i>H-A-<math>\lambda</math>-1</i> [27]	140.32	1.6583
5) <i>H-A-<math>\lambda</math>-2</i> [16]	138.93	5.2731
6) <i>GP2HPI</i> [15]	<b>100.56</b>	<b>1.0979</b>

The Global BnB was computationally expensive for each new position, obtaining an average of 223.9 seconds (the maximum value registered was 546.2 sec and the minimum was 3.25 sec) for 100 random poses. However, all 100 images converged to the true pose. The most computationally demanding part was the exploration of sub-cubes in the BnB algorithm and the time shown in Table II is the elapsed time of the local method to get convergence after the BnB initialization (non cumulative). It can be noted that many of the compared strategies are improved since the BnB gives a rough pre-alignment for these local approaches.

*Experiment 2:* Well known RGB-D Benchmarks [9] were used to perform keyframe visual odometry. Keyframes are manually selected from the sequences at every  $n$ -th frame. The local methods initialize the registration. If the local method cannot obtain an optimal convergence, then it is initialized with the BnB algorithm. In Table I, the Absolute Trajectory Error and Relative Pose Error evaluation are shown. They demonstrate that the initialization with BnB improve the estimations. Less error is obtained w.r.t. other

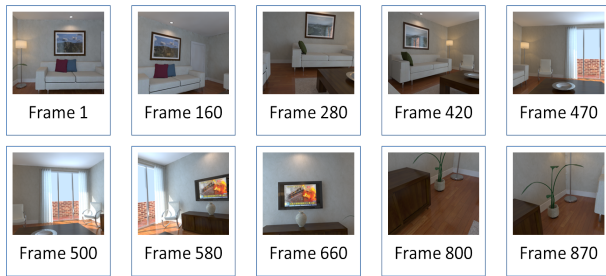
local classic approaches by combining the local Point-to-hyperplane ICP, which reduce the computational cost. The time shown does not consider the computation of the normals or the kd-tree, but time for convergence. It can be observed that the local method can get trapped in local minima during the registration process (less number of times for the Point-to-hyperplane), which is avoided by calling the BnB method to initialize the pose. Since less frames are employed for the 3D visual odometry, the representation of the environment can be done by considering keyframes only.

*Experiment 3:* Pre-selected RGB-D frames were employed to generate 3D maps of closed loop sequences. The keyframes have a small overlap between them and they are globally aligned by using the Point-to-hyperplane strategy. The method performs RGB-D registration by following the sequence has in Experiment 2, but the alignment w.r.t. the generated model is employed if the optimal solution is not found. This can be referred as a frame-to-model registration with the advantage that the model is generated simultaneously. Normally, a frame cannot be aligned when similar geometric and photometric properties are present in the overlapping area. When the entire generated cloud is considered, more correspondences can be found by using the extended measurements (color + depth).

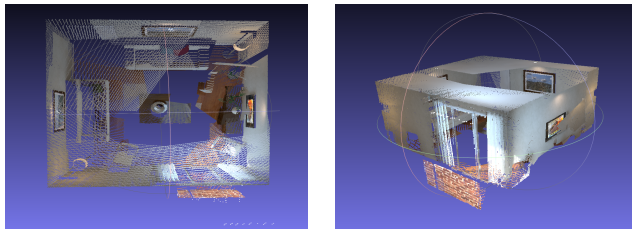
## V. CONCLUSIONS

In this paper, a global Point-to-hyperplane ICP method was proposed. The method was employed to improve RGB-D registration by optimally minimizing over fused color and depth measurements. The method combined the local Point-to-hyperplane approach with a global Branch and Bound method. The method is shown to handle the local minima problem. The extension to 4 dimensions was considered and it was noted that global methods based on an uncertainty radius cannot be easily extended to higher dimensions due to the fact that a  $n$ -sphere decreases its volume after dimension 5. However, the method demonstrated an improved performance by considering 4 dimensions. As a future work, more strategies to compute bounds will be explored and a selection of keyframes strategy will be implemented.



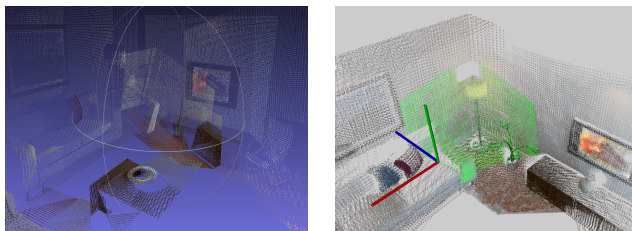


(a) RGB-D keyframes



(b) top view

(c) perspective view



(d) error

(e) fixed error

Fig. 7. Example of a 3D reconstruction (b) and (c) by considering only pre-selected keyframes (a). The proposed method obtains close solutions w.r.t. groundtruth. (d) The last image (frame 870) generates an error (d) since it contains similar geometric and color features w.r.t. the previous frame (800) but it can be fixed if the entire generated point cloud is considered for global registration (e).

## REFERENCES

- [1] P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(2):239–256, Feb 1992.
- [2] A. Parra Bustos, T. J. Chin, A. Eriksson, H. Li, and D. Suter. Fast rotation search with stereographic projections for 3d registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2227–2240, Nov 2016.
- [3] D. Campbell and L. Petersson. Gogma: Globally-optimal gaussian mixture alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5685–5694, June 2016.
- [4] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image Vision Comput.*, 10(3):145–155, April 1992.
- [5] Andrew I Comport, Ezio Malis, and Patrick Rives. Accurate quadrifocal tracking for robust 3d visual odometry. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 40–45. IEEE, 2007.
- [6] Andrew Fitzgibbon. Robust registration of 2d and 3d point sets. *Image and Vision Computing*, 21/14:1145–1153, January 2003.
- [7] Natasha Gelfand, Niloy J. Mitra, Leonidas J. Guibas, and Helmut Pottmann. Robust global registration. In *Proceedings of the Third Eurographics Symposium on Geometry Processing*, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.
- [8] Maciej Halber and Thomas A. Funkhouser. Structured global registration of RGB-D scans in indoor environments. 2016.
- [9] A. Handa, T. Whelan, J. McDonald, and A.J. Davison. A benchmark for rgb-d visual odometry, 3d reconstruction and slam. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 1524–1531, May 2014.
- [10] Richard I. Hartley and Fredrik Kahl. Global optimization through rotation space search. *International Journal of Computer Vision*, 82(1):64–79, Apr 2009.
- [11] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, Apr 1987.
- [12] Daniel Huber and Martial Hebert. Fully automatic registration of multiple 3d data sets. 21(7):637–650, July 2003.
- [13] P.J. Huber, J. Wiley, and W. InterScience. *Robust statistics*. Wiley New York, 1981.
- [14] Fernando I. Ireta Muñoz and Andrew I. Comport. A proof that fusing measurements using point-to-hyperplane registration is invariant to relative scale. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Baden-Baden, Germany, 2016.
- [15] Fernando I. Ireta Muñoz and Andrew I. Comport. Point-to-hyperplane RGB-D Pose Estimation: Fusing Photometric and Geometric Measurements. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Daejeon, South Korea, October 2016.
- [16] C. Kerl, J. Sturm, and D. Cremers. Dense visual slam for rgb-d cameras. In *International Conference on Intelligent Robot and Systems*, 2013.
- [17] E. L. Lawler and D. E. Wood. Branch-and-bound methods: A survey. *Operations Research*, 14(4):699–719, 1966.
- [18] A. Makadia, A. Patterson, and K. Daniilidis. Fully automatic registration of 3d point clouds. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1297–1304, June 2006.
- [19] M. Meilland and A.I. Comport. On unifying key-frame and voxel-based dense visual SLAM at large scales. In *International Conference on Intelligent Robots and Systems*, Tokyo, Japan, 3-8 November 2013. IEEE/RSJ.
- [20] Nicolas Mellado, Dror Aiger, and Niloy J. Mitra. Super 4pcs fast global pointcloud registration via smart indexing. *Computer Graphics Forum*, 33(5):205–215, 2014.
- [21] Marius Muja and David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Application*, pages 331–340. INSTICC Press, 2009.
- [22] C. Olsson, F. Kahl, and M. Oskarsson. Branch-and-bound methods for euclidean registration problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):783–794, May 2009.
- [23] Harold R. Parks. The volume of the unit n-ball. *Mathematics Magazine*, 86(4):270–274, 2013.
- [24] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the ICP algorithm. In *Third International Conference on 3D Digital Imaging and Modeling*, June 2001.
- [25] A. Segal, D. Haehnel, and S. Thrun. Generalized-icp. In *Proceedings of Robotics: Science and Systems*, Seattle, USA, June 2009.
- [26] Julian Straub, Trevor Campbell, Jonathan P. How, and John W. Fisher III. Efficient globally optimal point cloud alignment using bayesian nonparametric mixtures. 2016.
- [27] T.M. Tykkälä, C. Audras, and A.I. Comport. Direct Iterative Closest Point for Real-time Visual Odometry. In *The Second international Workshop on Computer Vision in Vehicle Technology: From Earth to Mars in conjunction with the International Conference on Computer Vision*, Barcelona, Spain, November 6-13 2011.
- [28] Jiaolong Yang, Hongdong Li, and Yunde Jia. Go-icp: Solving 3d registration efficiently and globally optimally. In *Proceedings of the 2013 IEEE International Conference on Computer Vision*, pages 1457–1464, Washington, DC, USA, 2013. IEEE Computer Society.
- [29] Yu Zhang Zheng Fang. Experimental evaluation of rgb-d visual odometry methods. *International Journal of Advanced Robotic Systems*, 2015-03-27.
- [30] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. *Fast Global Registration*, pages 766–782. Springer International Publishing, Cham, 2016.