PHASER: A Robust and Correspondence-Free Global Pointcloud Registration

Lukas Bernreiter [ID], Lionel Ott [ID], Juan Nieto [ID], Roland Siegwart [ID], and Cesar Cadena [ID]

Abstract—We propose PHASER, a correspondence-free global registration of sensor-centric pointclouds that is robust to noise, sparsity, and partial overlaps. Our method can seamlessly handle multimodal information, and does not rely on keypoint nor descriptor preprocessing modules. By exploiting properties of Fourier analysis, PHASER operates directly on the sensor's signal, fusing the spectra of multiple channels and computing the 6-DoF transformation based on correlation. Our registration pipeline starts by finding the most likely rotation $r \in SO(3)$ followed by computing the most likely translation $t \in \mathbb{R}^3$. Both estimates, r, and t are distributed according to a probability distribution that takes the underlying manifold into account, i.e., a Bingham and a Gaussian distribution, respectively. This further allows our approach to consider the periodic nature of r and naturally represents its uncertainty. We extensively compare PHASER against several well-known registration algorithms on both simulated datasets, and real-world data acquired using different sensor configurations. Our results show that PHASER can globally align pointclouds in less than 100 ms with an average accuracy of 2 cm and $0.5°$, is resilient against noise, and can handle partial overlap.

Index Terms—Mapping, probability and statistical methods, sensor fusion.

I. INTRODUCTION

THE registration of pointclouds from 3D scanners stems as a core competence for numerous applications in robotics such as state estimation, map building, and localization [1], [2]. Especially since recent advances of more affordable and effective RGB-D and LiDAR scanners leverage to infer an environment's geometry readily. It is further particularly crucial for ill-lighted and unstructured environments, where LiDAR systems constitute the preferred sensing option, as they are inherently invariant to illumination changes yielding a potentially more reliable robotics system.

Traditional local approaches such as Iterative Closest Point (ICP) are highly sensitive to local minima in the objective function, making a good prior initialization of the registration

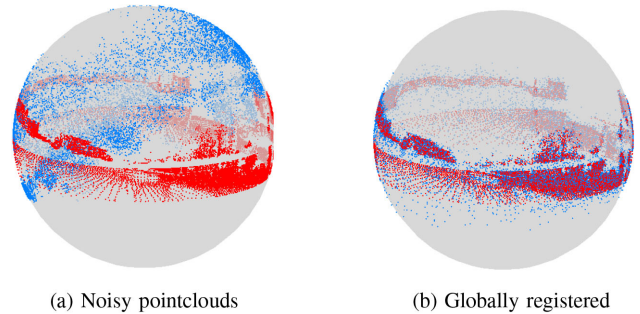(a) Noisy pointclouds   (b) Globally registered

Fig. 1. We propose a global registration pipeline by employing a Fourier analysis on the pointclouds. Given two pointclouds in (a), we find the correct global alignment for the rotation and translation, as in (b).

an indispensable requirement. However, often in robotics, only a very poor initialization is available. Therefore, global registration systems have gained increased interest since they can provide an optimal solution without requiring an initial guess. Many of the existing global registration approaches rely on computing point correspondences by matching unique point descriptors between the scans [3] stating that a subset of points from the source scan corresponds to another in the target scan. However, relying on such point descriptors raises the concern that an algorithm's efficiency and success depend highly on the extracted descriptor's quality. Especially in the context of localization, where the goal is to find a transformation between a local scan to a prior map, this is inherently flawed since the points are distorted and noisy. Hence, they perturb the extraction of the descriptors and consequently the matching. Furthermore, the global alignment problem becomes significantly more challenging when the data contains different viewpoints yielding confined overlap between the pointclouds and many outlier correspondences. Alternatively, approaches which do not rely on correspondences are less affected by such issues and can potentially increase the overall system's robustness [4].

Also, it is evident that in many of the aforementioned cases, an assessment of the alignment quality is another critical aspect for several robotic applications, for instance, its uncertainty for incorporating loop-closure factors in graph-based SLAM. However, finding the proper balance between the computation time and the uncertainty's precision is considerably challenging and remains an active research field [5].

Moreover, conventional global pointcloud registration systems often neglect the additional channels and information a LiDAR sensor provides. Besides range information, LiDAR sensors typically also return the received energy level (intensity)
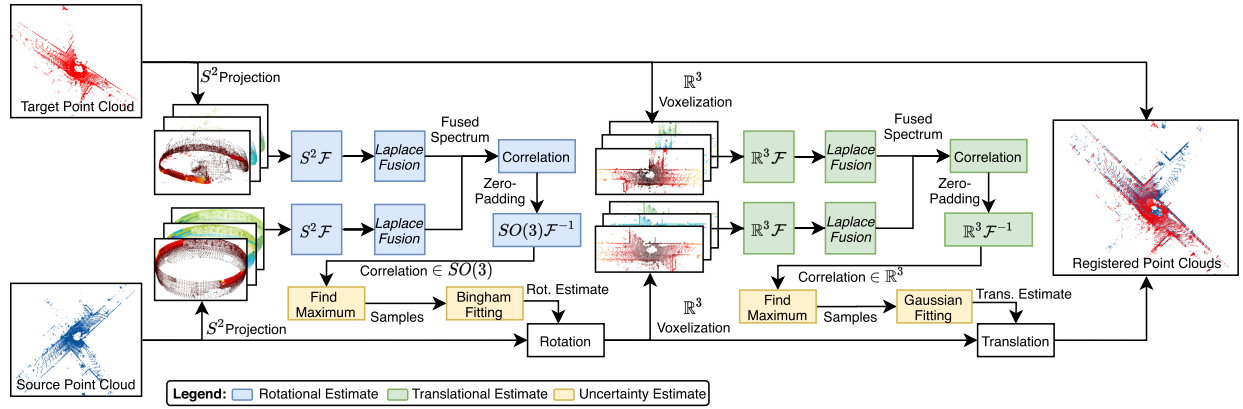
Fig. 2. Overview of the proposed system. For the registration core, we project both, source and target pointcloud onto the sphere and perform a spherical Fourier transform ($S^2\mathcal{F}$) to find the rotation that maximizes the spherical correlation (blue). Using the oriented scans, we estimate the translation as a next step using a spatial Fourier transform ($\mathbb{R}^3\mathcal{F}$) and correlation (green). Our framework additionally examines each correlation's result to infer a measure of the registration uncertainty (yellow). If additional modalities are used, our pipeline will automatically fuse their spectra using a Laplacian pyramid before the correlation step is performed.

and the ambient IR. Both modalities represent additional characteristics of the reflected surface and provide valuable information about the environment but need careful consideration when incorporated into a registration pipeline.

In this letter we propose a solution based on Fourier analysis. The approach transforms the pointclouds to the frequency domain, where the frequency components of multiple modalities are fused and then aligned through correlation. As a consequence, our approach does not require any point correspondences, initial guess, nor additional parameters for the fusion. We make use of a spherical projection model (cf. Figure 1) to enable a decomposition into spherical frequencies of the pointclouds. Furthermore, we derive a measure for the registration uncertainty using the outcome of the correlation. In more detail, for the translation we fit a multivariate Gaussian distribution to the resulting correlation to estimate the spatial uncertainty. For the rotation, we make use of directional statistics to account for the fact that the estimate does not lie in the Euclidean space but rather on a curved manifold. Specifically, we use the Bingham distribution which inherently takes the periodic nature of the rotation into account. A general overview of our registration pipeline is given in Figure 2. We denote our proposed registration pipeline as *PHASER*, short for phase registration.

The resulting global registration pipeline is evaluated in simulations where we can verify the correctness of the algorithm, and in real-world experiments to show its pragmatic applicability. The main contributions of this work are:

- A spherical and spatial Fourier decomposition of band-limited signals to individually estimate the rotation and translation of pointclouds.
- A fusion methodology for arbitrary input modalities.
- A method for estimation of the spherical and spatial correlation uncertainty using a Bingham and a Gaussian distribution, respectively.

II. RELATED WORK

Pointcloud registration is a well studied field with a variety of different applications and there is extensive literature available,

e.g. Pomerleau [2] provides a good overview in the context of robotics. Typically, registration approaches can be divided into local and global approaches and further can rely on correspondences between the pointclouds or operate correspondence-free.

A. Local Pointcloud Registration

If the correspondences are known it was shown in the seminal work of Horn [6] that registration can be solved in closed form. For unknown correspondences, traditional approaches such as ICP iteratively refine the correspondences and compute the optimal solution using Horn's method [6] at each iteration. Consequently, the registration quality substantially decreases in the presence of noise and motion-distortions. Therefore, local methods as ICP strongly depend on a good initial prior. To overcome some robustness issues, probabilistic approaches such as NDT [7], Generalized-ICP [8] and the GMM-based registration [9] fit a Gaussian distribution to the points or voxels and optimize over a distribution-to-distribution distance.

While the majority relies on geometric information, there has been some work which includes additional information such as color or intensity to e.g. reject outliers [10] or partition distributions [11]. Compared to all previously mentioned methods, ours does not minimize a metrical error in vector space but rather measures the similarity between clouds and therefore, operates globally.

B. Global Pointcloud Registration

Global pointcloud registration approaches, such as the work of Rusu *et al.* [3] and Lei *et al.* [12] seek to efficiently find a unique solution by extracting point descriptors to find correspondences for the registration without prior knowledge. The work Le *et al.* [4] proposes a randomized approach to overcome issues with computing descriptors and point correspondences. Yang *et al.* [13] proposed a global variant of the ICP algorithm based on the branch-and-bound technique for global optimization.

Yang and Carlone [14] approach the motion estimation by decoupling the rotation from the translation estimate. Their method targets the case of extreme outlier correspondences and

shows increased robustness against these. Making use of the Fourier domain also enables decoupling rotation from translation which is based on the fact that phase shifts do not affect the magnitude of a transformed signal [15], [16]. Correlation-based registration approaches in the Fourier domain have shown good robustness for finding translations in 2D [17] and 3D [18] and further, do not require computing correspondences. Bülow *et al.* [15] propose a heuristic-based resampling of FFT-layers to compute the rotational difference while the translational offset is found by matching the phases. Generalizing the Fourier transform to the spherical domain additionally enables finding the orientation between pointclouds [16], [19]. Bülow and Birk [16] proposed an extension to the well known Fourier-Mellin transform to include a spherical correlation resulting in a pipeline for estimating rotation, translation, and scale. Further, Makadia *et al.* proposed a full registration pipeline for pointclouds [19] utilizing orientation histograms of normals and a convolution over the occupancy. Likewise, our method's registration core also separates rotation from translation and transforms the input using a spherical and spatial Fourier transform, respectively. In contrast to [16], [19], our approach encodes more characteristics such as signal intensities instead of only occupancy information.

Many of the mentioned global pointcloud registration approaches are often tailored to 3D scanned objects with uniformly distributed points or to provide only a coarse alignment which is later fine-tuned by ICP. Notably, LiDAR scans in robotics are often distorted, have a large, irregular spatial extent, and are affected by noise. Thus, require a robust registration methodology and an evaluation of its quality. Our method can deal with high noise levels by operating in the spherical and spatial Fourier domain where we fuse and correlate the spectra of multiple modalities using a Laplacian pyramid. We can further configure our pipeline for a coarse or fine registration and, using the correlation, approximate the registration uncertainty.

C. Registration Uncertainty

Landry *et al.* [5]. present an overview of current approaches for uncertainty estimation for registration. The solutions can be grouped into two main groups: sampling-based approaches [20] and closed-form solutions [21]. Typically, sampling-based approaches have high computational demand and often do not scale well with large search spaces. In contrast, closed-form solutions are often less computational-intensive but should be used cautiously when noise is present [5]. The solution presented in this letter differs by requiring only a small amount of weighted correlation samples, which are then fitted to a probability distribution.

III. PRELIMINARIES

We address the registration problem by performing a correlation between source and target pointcloud in the Fourier domain. Thus, we first introduce the necessary concepts needed to perform a correlation in the spherical and spatial Fourier domain. This essentially covers the theory of harmonic analysis on the rotation group as well as on the vector space. Further, we introduce the directional distribution used for the orientation uncertainty.

A. Spatial and Spherical Fourier Analysis

In general, the spatial Fourier transform of a function $f \in \mathbb{R}^3$ is defined as the inner product $F(\boldsymbol{s}) = \langle f, \exp(j2\pi\boldsymbol{s}) \rangle$, where j is the imaginary unit and $\boldsymbol{s} = [u, v, w]^\top$ the spatial frequency coordinates. Using n^3 equally spaced samples (i.e. voxels) over $\mathbb{R}^3$ leads to its well known discrete form, i.e.

$$F(u, v, w) = \frac{1}{n^3} \sum_{x=0}^{n-1} \sum_{y=0}^{n-1} \sum_{z=0}^{n-1} f(x, y, z)$$
$$\times \exp\left(-j\frac{2\pi}{n}[x, y, z]^\top[u, v, w]\right). \tag{1}$$

Generally, given the sensors' field of view FoV, the spatial resolution $\Delta = \frac{FoV}{n^3}$ (i.e. voxel width) directly controls the used bandwidth $B = \frac{2\pi}{\Delta}$ in the frequency domain and consequently the discretization of the spectrum.

Moreover, it is well known that the Fourier coefficients retain their phase information under translation of the original signal. Thus, a translational shift in spatial coordinates results in a phase difference in the frequency domain which can be readily found using correlation techniques [17], [18].

Generally, the theory behind the Fourier analysis of signals can be extended to the hypersphere $S^2 = \{x \in \mathbb{R}^3 \mid \|x\|_2 = 1\}$, where we use the same parametrization as in Healy *et al.* [22], i.e. $\boldsymbol{\omega}(\phi, \theta) = [\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta]^\top$ for $\boldsymbol{\omega} \in S^2$, $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi]$. The Hilbert space of $L^2(S^2)$ is defined over the inner product as

$$\langle \tilde{f}, \tilde{h} \rangle = \int_{\boldsymbol{\omega} \in S^2} \tilde{f}(\boldsymbol{\omega})\overline{\tilde{h}(\boldsymbol{\omega})}d\boldsymbol{\omega}, \tag{2}$$

where $d\boldsymbol{\omega} = \sin(\theta)d\theta d\phi$, $\tilde{f}, \tilde{h}$ are arbitrary square-integrable functions on S^2 and $\overline{\tilde{h}}$ its complex conjugate.

The spherical Fourier transform of any function $\tilde{f} \in S^2$ is defined as the inner product $\tilde{F}_m^l = \langle \tilde{f}, Y_m^l \rangle$, where Y_m^l are the so called spherical harmonics of degree $l \in \mathbb{N}_0$, order $m \in [0, l] \in \mathbb{N}_0$ and form an orthonormal basis over $L^2(S^2)$, i.e.

$$Y_m^l(\boldsymbol{\omega}) = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_m^l(\cos\theta)\exp(-jm\phi), \tag{3}$$

where P_m^l are the associated Legendre polynomials which can be calculated numerically stable as proposed in the work by Healy *et al.* [23].

We evaluate the integral over the sphere in eq. 2 using equiangular samples according to the sampling theorem by Driscoll and Healy [24] which results in the following Fourier coefficients:

$$\tilde{F}_m^l = \frac{\sqrt{2\pi}}{2\tilde{B}} \sum_{j=0}^{2\tilde{B}-1} \sum_{k=0}^{2\tilde{B}-1} w_j \tilde{f}(\theta_j, \phi_k)\overline{Y}_m^l(\theta_j, \phi_k), \tag{4}$$

where $\tilde{B}$ is the spherical bandwidth, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are the samples with the corresponding weights $\boldsymbol{w}$. Combining eq. (4) to a set of coefficients results in the spherical Fourier transform for a given $\tilde{B}$.

Furthermore, this enables the expansion of arbitrary functions $\tilde{f} \in L^2(S^2)$ in the base of the spherical harmonics, i.e.

$$\tilde{f}(\boldsymbol{\omega}) = \sum_{l \geq 0} \sum_{m \leq l} \tilde{F}_m^l Y_m^l(\boldsymbol{\omega}). \tag{5}$$

Our orientation estimation for pointclouds is based on the fact that the magnitude $|\tilde{F}|$ is invariant under rotation [25]. Thus, similar to the case of a spatial displacement, the rotational difference between two signals can be found using correlation techniques [16], [19].

B. Orientation Uncertainty Representation

In this work we make use of the Bingham ditribution, a directional distribution, which naturally arises when conditioning a zero-mean Gaussian distribution on the hypersphere [26]. Its probability density function is defined as

$$\mathcal{B}(\boldsymbol{x}; M, Z) = \frac{1}{N(Z)} \exp\left(\boldsymbol{x}^\top M Z M^\top \boldsymbol{x}\right), \tag{6}$$

where M is the orientation, Z are the concentration parameters and $N(Z)$ a normalization constant.

The Bingham distribution is antipodal symmetric, takes the periodicity into account, and also naturally models the theory of unit-quaternions, i.e. the fact that $\boldsymbol{q}$ and $-\boldsymbol{q}$ represent the same rotation. Consequently, an uncertain orientation can be modeled as a unit-quaternion that is distributed according to a 4D Bingham distribution. With this formulation we can now utilize the second-order moment of a Bingham distribution as an uncertainty metric. It can be derived as follows

$$\Sigma_{\mathcal{B}} = M \cdot \operatorname{diag}\left(\frac{\frac{\partial}{\partial z_1} N(Z)}{N(Z)}, \ldots, \frac{\frac{\partial}{\partial z_d} N(Z)}{N(Z)}\right) M^\top, \tag{7}$$

where d is the dimension of the distribution.

IV. METHOD

We seek to find a rigid-body motion in $SE(3)$, with its uncertainty, which transforms a source to a target pointcloud. In this section, we outline the core components of our pipeline shown in Figure 2. Starting with the source and target clouds, this comprises: (i) input representation, (ii) orientation estimation, (iii) translation estimation and (iv) fusion of multiple modalities.

A. Representative Functions

The input pointclouds are represented by two functions, i.e., $f, h \mathbb{R}^3$ and $\tilde{f}, \tilde{h} \in S^2$. Throughout this letter, we use f and $\tilde{f}$ for the target, h and $\tilde{h}$ for the source pointcloud.

Furthermore, we define a set of K modalities suitable for correlation analysis, i.e. f_i, h_i, and $\tilde{f}_i, \tilde{h}_i$ for $i \in \{1, \ldots, K\}$. Generally, these functions can be chosen arbitrarily, although they ought to represent the environment somehow. Simple mappings of the individual channels (e.g., intensity, reflectivity), or handcrafted features (e.g., corner and surface features), are suitable choices. However, we will first cover the registration core with a single modality ($K = 1$) for simplicity and denote $f = f_i, h = h_i$ and $\tilde{f} = \tilde{f}_i, \tilde{h} = \tilde{h}_i$ before we discuss the fusion of different modalities. The functions f, h and $\tilde{f}, \tilde{h}$ are transformed, fused, and correlated in the spatial and spherical

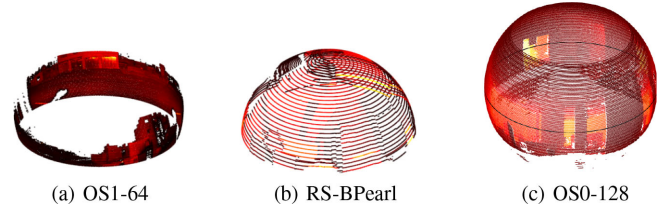(a) OS1-64   (b) RS-BPearl   (c) OS0-128

Fig. 3.   Illustration of the resulting spherical projections using different LiDAR sensors. Projection (a) shows the result of an Ouster OS1 with 64 beams, (b) a Robosense-BPearl with 32 beams, and (c) an Ouster OS0 with 128 beams.

frequency domain, respectively, to infer the relative transformation between them. Since the translational part requires oriented pointclouds, we first perform the spherical correlation.

B. Rotation Estimation

As an initial step, we project each pointcloud onto the sphere, i.e. the jth point $\boldsymbol{p}_j = [x_j, y_j, z_j]^\top$ is projected using

$$\phi_j = \arctan\left(\frac{y_j}{x_j}\right), \quad \theta_j = \arccos\left(\frac{z_j}{||\boldsymbol{p}_j||_2}\right), \tag{8}$$

where ϕ_j is the azimutal and θ_j the polar angle as defined in Section III-A. Figure 3 illustrates the projections of three different LiDAR sensors.

After the projection, we employ the sampling theorem by Driscoll and Healy [24] to create a $2\tilde{B} \times 2\tilde{B}$ sampling grid on the sphere. We employ a nearest neighbor search using the grid points to get the sampled pointcloud where we average points that fall within the same bin, s.t. this discretization yields $\tilde{f}, \tilde{h} \mathbb{R}^{2\tilde{B} \times 2\tilde{B} \times K}$. This enables the use of the spherical Fourier transform (cf. eq. (4)) yielding $\tilde{F}$ and $\tilde{H}$. The spherical correlation $\tilde{C}$ of two functions $\tilde{f}$ and $\tilde{h}$ is defined as their inner product, i.e. $\tilde{C} = \langle \tilde{f}, \tilde{h} \rangle$.

If $\tilde{h}$ is a rotated version of $\tilde{f}$, we find the rotation $\boldsymbol{r} \in SO(3)$ by maximizing $\tilde{C}$, i.e.

$$\underset{\boldsymbol{r} \in SO(3)}{\arg\max} \langle \tilde{f}, \boldsymbol{r}^{-1}\tilde{h} \rangle, \tag{9}$$

where the inner product is defined as in eq. (2). As proposed by Kostelec and Rockmore [25], by making use of eq. (5) and the fact that spherical harmonics are orthogonal, eq. (9) can be rewritten and efficiently evaluated using the spherical Fourier coefficients, i.e.

$$\tilde{C} = \sum_{l > 0} \sum_{m \leq l} \sum_{m' \leq l} \tilde{F}_m^l \overline{\tilde{H}_{m'}^l} D_{mm'}^l, \tag{10}$$

where D denotes the Wigner-D functions. For the reader interested in a detailed derivation of eq. (10) please refer to [25].

After correlating the coefficients, we zero-pad the coefficients vector to increase the resulting resolution and perform a subsequent inverse $SO(3)$ Fourier transform [25] to find the rotation $\boldsymbol{r}$ that maximizes eq. (9). Thereby, we find the maximum, which corresponds to our best estimate for the rotational alignment and its neighbors in the correlation tensor. In concrete, each signal in the correlation tensor is a rotation that is initially represented as ZYZ-Euler angles. Using the normalized magnitude as the weight, we convert the ZYZ-Euler angles to a unit-quaternion

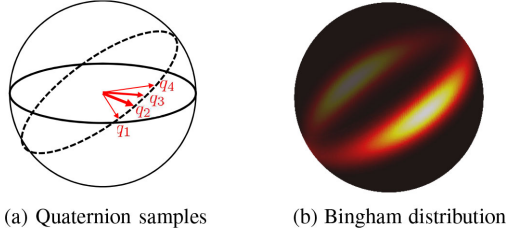(a) Quaternion samples   (b) Bingham distribution

Fig. 4. (a) Illustrates different rotational estimates on the unit sphere. The thickness of the arrows denotes the weight of the estimate. (b) Shows a fitted Bingham distribution from weighted samples. Generally, noise and partial overlap result in high weighted candidates and can perturb correlations. Our proposed fusion of multiple modalities alleviates the correlation bias and yields a more precise result.

and fit the maximum peak as well as its neighbors to a Bingham distribution (cf. Figure 4). In more detail, we perform a maximum likelihood estimation by matching the second-order moment with the unit-quaternion samples $Q_\mathcal{B}$ and their corresponding weights $w_\mathcal{B}$, i.e.

$$\Sigma_\mathcal{B} = Q_\mathcal{B} W_\mathcal{B} Q_\mathcal{B}^\top = U \Lambda U^\top, \qquad (11)$$

where $W_\mathcal{B} = \mathrm{diag}(w_\mathcal{B})$, and where $\Lambda = \mathrm{diag}(\lambda)$ and U are the eigenvalues and eigenvectors after an eigendecomposition of $\Sigma_\mathcal{B}$. Using the relationship in eq. (7), the eigenvectors U correspond to the orientation M, and λ_i corresponds to $\frac{\frac{\partial}{\partial z_i} N(Z)}{N(Z)}$. Finally, we use a local and convex optimization to derive the optimal concentration parameter $\hat{Z}$ using the following cost function

$$J(\hat{Z}) = \sqrt{\sum_i \left(\frac{\frac{\partial}{\partial \hat{z}_i} N(\hat{Z})}{N(\hat{Z})} - \lambda_i \right)^2}, \qquad (12)$$

where λ_i is the i-th normalized eigenvalue. The fitted Bingham distribution comprises the rotational estimation part of the registration. A limitation of our uncertainty derivation is that this does not capture the uncertainty of other potential peaks in the correlation but comprises only the highest correlating rotation.

At this point, we take one of the modes of Bingham and rotate the source pointcloud accordingly. Since the Bingham distribution is antipodal symmetric, it is indistinctive which mode we take.

C. Translation Estimation

After estimating the rotation, the source pointcloud is correctly oriented, enabling us to estimate the translation using a correlation for the full alignment. We first discretize the pointcloud space $\mathbb{R}^3$ in n equal dimensions with zero-padding to satisfy the requirements of eq. (1). The two discretized pointclouds $f, h \in \mathbb{R}^{n \times n \times n \times K}$ from Section IV-A are transformed to the spatial frequency domain using eq. (1), resulting in F and H, respectively. Similar to the rotational registration, we perform another correlation C of the coefficients to estimate the phase shift, i.e.

$$C = F \cdot \overline{H}, \qquad (13)$$

where $\overline{H}$ is the complex conjugate of H. Zero-padding as well as a subsequent inverse Fourier transform of C,

$$c(x, y, z) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \sum_{w=0}^{N-1} C(u, v, w)$$

$$\times \exp\left(j \frac{2\pi}{N} [x, y, z]^\top [u, v, w] \right), \qquad (14)$$

yields the final translation by finding the most prominent peaks of $c(\boldsymbol{x})$. In a similar fashion as in Section IV-B, we search for the maxima in the correlation function. Again, each signal defines a specific relative translation weighted by the magnitude, and the maximum corresponds to the best estimate of it. Using the maximum and its neighbors as samples T we fit a multivariate Gaussian distribution where the mean $\boldsymbol{\mu}$ is the weighted average and covariance $\Sigma_\mathcal{N}$ is defined as

$$\Sigma_\mathcal{N} = (T - \boldsymbol{\mu}) W_\mathcal{N} (T - \boldsymbol{\mu})^\top, \qquad (15)$$

where $W_\mathcal{N} = \mathrm{diag}(w_\mathcal{N})$ is a diagonal matrix of individual weights $w_\mathcal{N}$.

This results in our translational estimate, and therefore we have derived the relative transformation between the source and the target pointcloud for a single modality.

D. Multi-Modal Fusion

The previous sections cover the registration core of PHASER using a single modality. Given a set of representative functions (cf. Section IV-A) transformed in the spherical and spatial frequency domain, we now seek to find a fused spectrum for source and target pointcloud. Having a single fused spectrum allows us to leave the correlation pipeline, i.e., registration core of PHASER, to remain unchanged.

Spectral image fusion techniques [27] combine multiple complementary sensors to create an enhanced image by analyzing the spectral bands and selecting the most relevant parts of each sensor [28]. Inspired by these techniques, we employ a Laplacian pyramid-based fusion to construct a fused pointcloud spectrum from different modalities.

Initially, we construct a Gaussian pyramid where each level represents a low-pass filtered and subsampled by half version of the previous level. The difference of each level to the previous level denotes the Laplacian pyramid. At the first level, we utilize the difference to the original pointcloud spectrum. Range and intensity information yield a good spectral diversity, and when fused, provide a more enriched representation of the environment. Consequently, we construct a Laplacian pyramid for each input channel representing the details (contrast changes) for each channel. We fuse the channels by reconstructing a single merged pointcloud spectrum by combining the Laplacian pyramids to a single pyramid. Specifically, the top-level low-pass spectra are averaged, whereas all the other levels select the coefficient with the highest local energy [28]. Finally, the fused spectrum will be the input for the registration pipeline, as discussed in Section IV-B and Section IV-C.
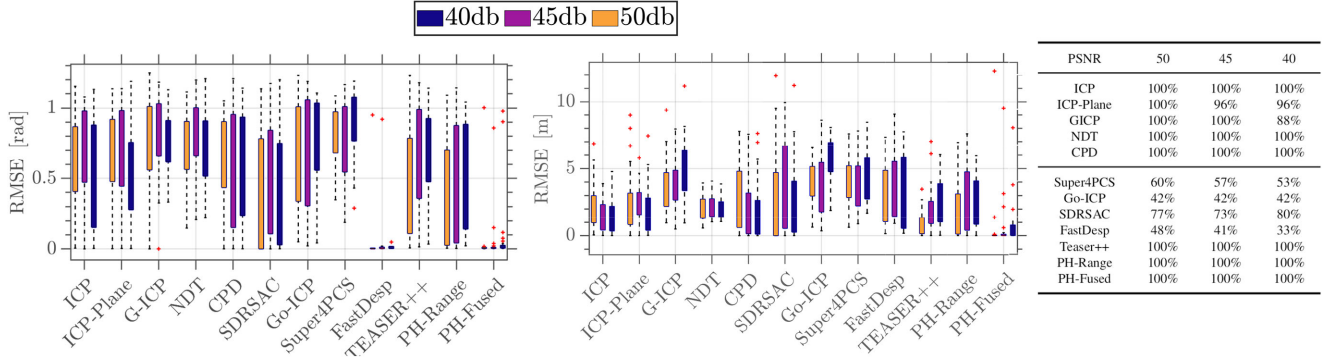
PSNR	50	45	40
ICP	100%	100%	100%
ICP-Plane	100%	96%	96%
GICP	100%	100%	88%
NDT	100%	100%	100%
CPD	100%	100%	100%
Super4PCS	60%	57%	53%
Go-ICP	42%	42%	42%
SDRSAC	77%	73%	80%
FastDesp	48%	41%	33%
Teaser++	100%	100%	100%
PH-Range	100%	100%	100%
PH-Fused	100%	100%	100%

Fig. 5. Evaluation of the RMSE in terms of rotation (left) and translation (middle) at different PSNR levels. We perform 80 random registrations for each PSNR level (240 in total) and show each method's success rate in the table (right). Here, *ICP-Plane* refers to the point-to-plane approach. *PH-Range* and *PH-Fused* denote PHASER using range-information only and a fusion of the available channels, respectively.

V. EXPERIMENTS

This section presents the experimental validation of our global alignment framework, where we show that our pipeline is robust against noise as well as partial overlap, converges to the global solution, and provides a meaningful uncertainty estimate. Throughout this section, we consider four local (ICP [29], NDT [7], CPD [30], G-ICP [8]) and five global (SDRSAC [4], FastDesp [12], Go-ICP [13], Super4PCS [31], TEASER++ [14]) registration methods for comparison. We used the respective authors' implementation except for ICP, NDT and CPD where we utilized the implementation that is available in MATLAB. For TEASER++, we first computed correspondences using FPFH [3] before estimating the relative transformations. We configure PHASER with $\tilde{B} = 120$, $B = 6$, i.e. $\tilde{f}, \tilde{h}\mathbb{R}^{240 \times 240 \times K}$ and $f, h\mathbb{R}^{200 \times 200 \times 200 \times K}$. Moreover, we use five pyramid levels and four neighbors for fitting the Bingham and Gaussian distribution. All experiments have been performed on real data and include a mix of indoor and outdoor environments.

A. Synthetic Benchmarks

In this experiment, we collected 60 pointclouds from robots equipped with a 64 beam Ouster OS-1 (65 535 points per cloud) and 20 pointclouds from the KITTI dataset [32] (Sequence 00) to simulate a global alignment problem. From the OS-1 dataset, we use four channels ($K = 4$): range, intensity, reflectivity, and ambient IR, whereas from the KITTI dataset, we utilize range and intensity ($K = 2$).

Noisy Measurements. We applied a random transformation to each pointcloud as well as additive zero-mean Gaussian noise to all of its channels. This results in a bijection between the transformed source and the initial target pointcloud. In particular, the experiments presented in this section aim to validate that our pipeline converges to the global solution independent of the initial alignment, i.e., it serves as a global registration method. Figure 5 shows the evaluation of the simulated transformations given different noise levels (cf. Figure 6) as well as their success rate. We consider an estimation unsuccessful if either the program crashes or the registration takes longer than 30 minutes.

Due to many local minima and limited search spaces, all local approaches cannot register the test cases correctly. The global
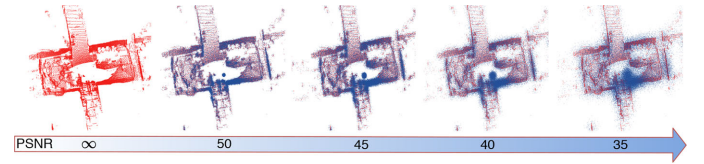

Fig. 6. Illustration of different PSNR levels. Red denotes the original pointcloud, blue the noised one. Our proposed pipeline is capable of robustly aligning pointclouds up to 40 dB PSNR.

search of Go-ICP provides a more extensive search space, but its least square objective is erroneous under the presence of noise. Super4PCS fails for all cases to correctly estimate the transformation since the noise disables the congruence search in the target pointcloud. SDRSAC showed better robustness due to its randomized approach yielding a broad error distribution. FastDesp and TEASER++, show competitive results despite being highly influenced by the noise during the descriptor extraction. Although our method is correspondence-free, it is similarly affected by the noise if we only consider the range information (*PH-Range*). However, if we supply our pipeline with all channels, we show that the registration is greatly improved (*PH-Fused*) even in high noise cases (40 dB) where we achieve an improvement of 91% and 73% in terms of the rotational and translation RMSE, respectively. Additionally, we achieve a similar rotational RMSE compared to FastDesp but gain a better success rate and an improvement of 80% for the 40 dB translational estimation. For TEASER++, we yield a 93% and 80% improvement w.r.t. to the 40 dB rotational and translational RMSE, respectively.

Missing Data. Next, we remove the artificial noise from the experiments and investigate the low overlapping case. Removing the noise does not corrupt the descriptor generation anymore, but instead, with little overlap, we provide less information to it. Consequently, we propose two experimental setups. First, we sparsified the pointclouds by randomly removing points from the source cloud before applying a random transformation. Second, given a percentage of points, we removed slices starting from the maximum towards the minimum resulting in pointclouds that only partially observed an environment, i.e., a simulated
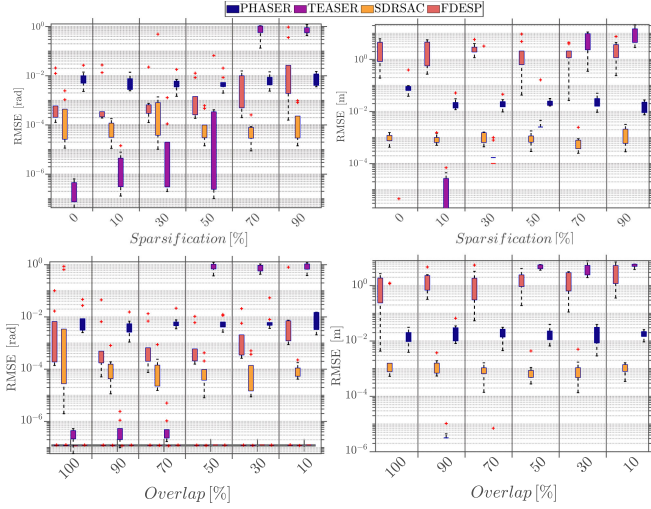
Fig. 7. Distribution of the RMSE of rotation (left) and translation (right) w.r.t. simulated sparsification (top) and simulated overlap (bottom) in log scale. Results are shown over 30 pointclouds registrations.
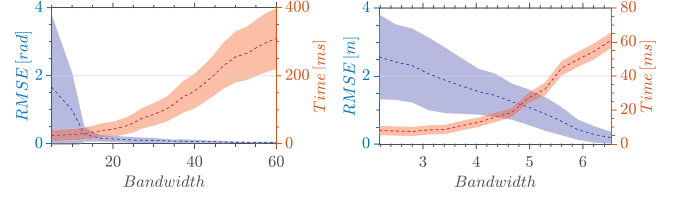

Fig. 8. Rotational (left) and translational (right) error w.r.t. computation time for a single modality. The execution time incorporates the discretization on S^2 and $hbbR^3$ as well as the individual correlations.
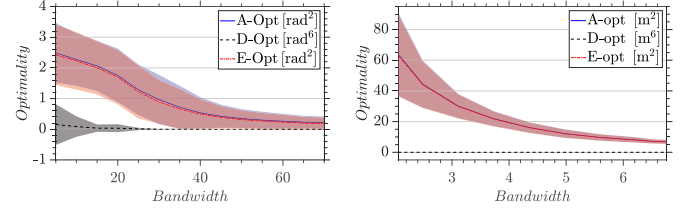

Fig. 9. Evaluation of the rotational covariance $\Sigma_{\alpha\beta\gamma}$ (left) and the positional covariance $\Sigma_{\mathcal{N}}$ (right) for different bandwidths. The line corresponds to the mean, and the shaded area denotes the standard deviation for each criterion.

partial overlap. Here, we further investigate the performance of SDRSAC, TEASER++ and FastDesp compared to PHASER since those yielded competitive results in the first experiment. Figure 7 illustrates the evaluation given different sparsification rates and overlap (partial observability) in log scale. In this experiment, SDRSAC attains significantly better registration quality since its graph matching performs better on noiseless data. For TEASER++, the RMSE starts low with little sparsification and high overlap and increases with large sparsification and little overlap. Removing large portions of the pointclouds leads to less precise descriptors for the source pointcloud resulting in little inlier rates. The multiscale descriptor approach of FastDesp yielded good rotational registration but failed to infer the translation for all of our datasets accurately. PHASER consistently estimates the transformation even at 90% sparsification and 10% overlap and maintains an RMSE of $0.4°$ and 0.15 m as we do not rely on any correspondences but rather measure the similarity of source and target pointcloud. Concluding, our approach enables accurate global registrations with noise and partial overlap. Our precision directly links to the spherical and spatial bandwidth and its resulting discretization, which we discuss in the next section.

B. Performance Evaluation

PHASER contains two main parameters, namely the spherical ($\tilde{B}$) and spatial (B) bandwidth. Both parameters control the registration pipeline, e.g., to have a sparse or dense sampling of the space. Figure 8 shows the trade-off between execution time and alignment error using different bandwidths. All experiments were executed on a laptop computer with an Intel i7-8850H. To summarize, our global alignment algorithm can be tailored to a coarse or fine registration by varying $\tilde{B}$ and B. Our experiments show that a small spherical bandwidth (20–30) already yields a reasonable rotational estimate and, together with a spatial bandwidth of 6-7 (cf. Figure 8), provides a good global alignment in less than 100 ms.

C. Uncertainty Analysis

In this section, we evaluate our uncertainty estimate based on three criteria [33]: (i) the trace of the covariance (A-opt), (ii) the determinant of the covariance (D-opt), and (iii) the maximum eigenvalue of the covariance (E-opt). Figure 9 shows the evaluation of these criteria, based on the dataset used in Section V-A with $PSNR = 50$ dB. These results infer that the uncertainty is consistent with the registration error (cf. Figure 8), i.e., it decreases with increasing bandwidth and hence with decreasing error.

Note that for better visualization of the rotational uncertainty, we created a sample-based Gaussian covariance of the half-sphere and converted it to an Euler angle covariance, $\Sigma_{\alpha\beta\gamma}$, using the covariance laws defined by Vanicek and Krakiwsky [34].

D. Real World Experiments

We recorded several runs in a search and rescue testing facility using a handheld device with an RTK GPS and a 128 beam Ouster OS-0 (131 072 points per scan). The environment comprises urban-like streets, buildings, and collapsed structures. In this experiment, we only utilize the range and intensity channels ($K = 2$). Given the GPS map (10 km long) with poses and scans, we randomly sampled inter and intra run pairs based on their positional proximity to evaluate how the distance affects our registration pipeline. As ground truth, we considered the relative GPS poses refined by ICP and grouped them by their relative distance. Specifically, in an interval of 0.25 m, we averaged over 20 registration each (420 registrations in total). Figure 10 illustrates the improvement of our proposed Laplacian-based fusion compared to a registration using range information only, TEASER++ and FastDesp. As expected, the registration using range information achieves a broad error distribution over the different distances while our proposed fusion stays mostly concise with an average RMSE of $6°$ and 0.7 m. Furthermore, we show that especially at greater distances (4–5 m) TEASER++
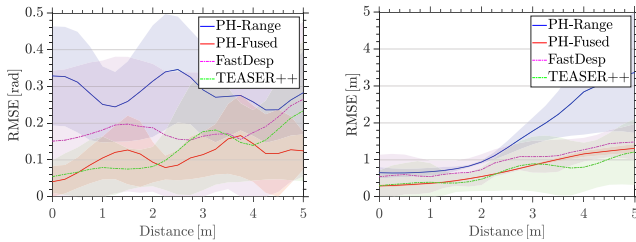
Fig. 10. Rotational (left) and translational (right) error w.r.t. distance between the poses. Utilizing a fusion of the range values with the intensity channel allows our pipeline to automatically choose the better modality for the registration and decrease the error accordingly.

and FastDesp have increased rotational error due to bad correspondences, whereas PHASER still maintains a low error.

VI. CONCLUSION AND FUTURE WORK

We have presented PHASER, a new framework for global pointcloud alignment based on the correlation of Fourier transformed signals. We make use of translational and rotational invariance of the spatial and spherical Fourier transform to estimate the relative transformation between pointclouds efficiently. Due to representation in the frequency domain, PHASER enables the use of multiple input modalities.

Our evaluations show that PHASER is robust to noisy measurements, different misalignment levels, and partial overlap. Thus, PHASER is shown to converge to the global solution, with better performance than previous methods. These findings can have relevant implications in scenarios such as multi-robot mapping or the kidnapped robot problem. Furthermore, we showed that our uncertainty estimate is consistent with the error of the alignment.

As part of our future work, we would like to find synergies between the frequencies in vector and spherical space and improve our uncertainty representation by considering multiple peaks in the correlation. Additionally, we would like to explore semantic information as a discrete random variable for the registration.

REFERENCES

[1] J. Zhang and S. Singh, "LOAM: LiDar odometry and mapping in real-time," *Robot. Sci. Syst.*, vol. 2, no. 9, 2014.

[2] F. Pomerleau, F. Colas, and R. Siegwart, "A Review of Point Cloud Registration Algorithms for Mobile Robotics," *Found. Trends Robot.*, vol. 4, no. 1, pp. 1–104, 2015.

[3] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc IEEE Int. Conf. Robot. Autom.*, 2009, pp. 3212–3217.

[4] H. Le, T.-T. Do, T. Hoang, and N.-M. Cheung, "SDRSAC: Semidefinite-based randomized approach for robust point cloud registration without correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 124–133.

[5] D. Landry, F. Pomerleau, and P. Giguère, "CELLO-3D: Estimating the covariance of ICP in the real world," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2019, pp. 8190–8196.

[6] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer.*, vol. 4, no. 4, pp. 629–642, 1987.

[7] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT," *J. Field Robot.*, vol. 24, no. 10, pp. 803–827, 2007.

[8] A. V. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," *Proc. Robot. Sci. Syst.*, vol. 2, p. 435, 2009.

[9] W. Tabib, C. O Meadhra, and N. Michael, "On-manifold GMM registration," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3805–3812, Oct. 2018.

[10] S. A. Parkison, M. Ghaffari, L. Gan, R. Zhang, A. K. Ushani, and R. M. Eustice, "Boosting shape registration algorithms via reproducing kernel Hilbert space regularizers," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 4563–4570, Oct. 2019.

[11] A. Zaganidis, L. Sun, T. Duckett, and G. Cielniak, "Integrating deep semantic segmentation into 3-D point cloud registration," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2942–2949, Oct. 2018.

[12] H. Lei, G. Jiang, and L. Quan, "Fast descriptors and correspondence propagation for robust global point cloud registration," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3614–3623, Aug. 2017.

[13] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016.

[14] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, 2020.

[15] H. Bülow and A. Birk, "Spectral 6DOF registration of noisy 3D range data with partial overlap," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 954–969, Apr. 2013.

[16] H. Bülow and A. Birk, "Scale-free registrations in 3D: 7 degrees of freedom with fourier mellin SOFT transforms," *Int. J. Comput. Vis.*, vol. 126, no. 7, pp. 731–750, 7 2018.

[17] B. Srinivasa Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

[18] C. Wang, X. Jing, and C. Zhao, "Local upsampling fourier transform for accurate 2D/3D image registration," *Comput. Elect. Eng.*, vol. 38, no. 5, pp. 1346–1357, 2012.

[19] A. Makadia, A. Patterson IV, and K. Daniilidis, "Fully automatic registration of 3D point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 1297–1304.

[20] J. Nieto, T. Bailey, and E. Nebot, "Scan-SLAM: Combining EKF-SLAM and Scan Correlation," in *Field and Service Robotics*. Berlin, Germany: Springer, 2006, vol. 25, pp. 167–178.

[21] S. M. Prakhya, L. Bingbing, Y. Rui, and W. Lin, "A closed-form estimate of 3D ICP covariance," in *Proc. 14th IAPR Int. Conf. Mach. Vis. Appl.*, 2015, pp. 526–529.

[22] D. M. Healy, D. N. Rockmore, P. J. Kostelec, and S. Moore, "FFTs for the 2-Sphere-Improvements and variations," *J. Fourier Anal. Appl.*, vol. 9, no. 4, pp. 341–385, 2003.

[23] D. M. Healy, D. N. Rockmore, and S. B. Moore, "FFT for the 2-sphere and applications," in *Proc. ICASSP*, *IEEE Int. Conf. Acoust., Speech Signal Process.*, 1996, pp. 1323–1326.

[24] J. R. Driscoll and D. M. Healy, "Computing Fourier Transforms and Convolutions on the 2-Sphere," *Adv. Appl. Math.*, vol. 15, no. 2, pp. 202–250, 1994.

[25] P. J. Kostelec and D. N. Rockmore, "FFTs on the rotation group," *J. Fourier Anal. Appl.*, vol. 14, no. 2, pp. 145–179, 2008.

[26] C. Bingham, "An antipodally symmetric distribution on the sphere," *Ann. Statist.*, pp. 1201–1225, 1974.

[27] V. P. Naidu, "Multi-resolution image fusion by FFT," in *Proc. Int. Conf. Image Inf. Process.*, 2011, pp. 1–6.

[28] H. Falk, "Prolog to a. categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proc. IEEE Proc. IRE*, vol. 87, no. 8, pp. 1313–1314, 1999.

[29] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms Data Structures*, vol. 1611. Int. Soc. Opt. Photon., Apr. 1992, pp. 586–606.

[30] A. Myronenko and X. Song, "Point set registration: Coherent point drifts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.

[31] N. Mellado, D. Aiger, and N. J. Mitra, "Super 4PCS fast global pointcloud registration via smart indexing," *Comput. Graph. Forum*, vol. 33, no. 5, pp. 205–215, 2014.

[32] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.

[33] H. Carrillo, I. Reid, and J. A. Castellanos, "On the comparison of uncertainty criteria for active SLAM," in *Proc," IEEE Int. Conf. Robot. Autom.*, 2012, pp. 2080–2087.

[34] P. Vanicek and E. J. Krakiwsky, *Geodesy: The Concepts*. New York, NY, USA: Elsevier Science, 2015.