Intelligent Service Robotics (2024) 17:1109–1124
https://doi.org/10.1007/s11370-024-00562-1

ORIGINAL RESEARCH PAPER

Robust and correspondence-free point cloud registration: an extended approach with multiple hypotheses evaluation

Federica Di Lauro[1] · **Domenico Giorgio Sorrenti**[1] · **Simone Fontana**[1]

Received: 15 December 2023 / Accepted: 18 September 2024 / Published online: 14 October 2024

Abstract
Point cloud registration is a fundamental problem in robotics, critical for tasks such as localization and mapping. Most approaches to this problem use feature-based techniques. However, these approaches have issues when dealing with unstructured environments where meaningful features are difficult to extract. Recently, an innovative global point cloud registration algorithm, PHASER, which does not rely on geometric features or point correspondences, has been introduced. It leverages Fourier transforms to identify the optimal rigid transform that maximizes cross-correlation between source and target point clouds. PHASER can also incorporate additional data channels, like LiDAR intensity, to enhance registration results. Because it does not rely on local features and because of its ability to exploit additional data, PHASER is particularly useful when dealing with very noisy point clouds or with many outliers. For this reason, we propose an extension to PHASER that considers multiple plausible rototranslation hypotheses. Our extended approach outperforms the original PHASER algorithm, especially in challenging scenarios where point clouds are widely separated. We validate its effectiveness on the DARPA SubT, and the Newer College datasets, showcasing its potential for improving registration accuracy in complex environments.

Keywords Point cloud · Registration · Slam · Point set · Robotic

1 Introduction

Point clouds registration, that is, finding the rigid transformation that best aligns two point clouds, is a very common problem in robotics. Indeed, it is a very important step for several robotic tasks, such as localization or mapping. There are two variants of this problem. The first one, called "local registration," employs local optimization techniques and works only when the two point clouds are already roughly aligned, that is, it is used to refine an alignment. A popular solution to this problem is ICP (Iterative Closest Point), developed independently by [1–3].

The second variant is called "global registration" and aims to align point clouds without any requirements on their initial alignment. Most solutions to this problems employ some kinds of local feature, which are used to represent the geometric structure of the environment. Examples are PFH [4] and its faster variant FPFH [5], Angle-Invariant Features [6], or SIFT features extracted from depth images [7].

Recently, feature extractors based on neural networks have been proposed too. One such method is fully convolutional geometric features, presented by Choy et al. [8]. This approach employs a fully convolutional neural network to produce geometric features in a single pass. Another notable example is 3D Match, a 3D convolutional neural network introduced by Zeng et al. [9]. 3D Match utilizes a 3D patch around a point to compute a feature descriptor. The distance between two descriptors determines the likelihood of a match. The descriptor is trained unsupervisedly with correspondences from existing RGB-D reconstructions. A different approach, 3DFeat-Net [10], relies on a Siamese architecture to learn to extract descriptive features, and on PointNet, introduced by Qi et al. [11], to represent point clouds. This model is trained to recognize whether two given point clouds originate from the same location, using weakly supervised learning with pairs of point clouds annotated with GPS and inertial-based localization. Unlike 3DMatch, 3DFeat-Net combines both a feature detector and a descriptor extractor. PPFNet [12] represents another approach that builds on PointNet. It utilizes points, normals, and point pair

✉ Simone Fontana
simone.fontana@unimib.it

[1] University of Milan-Bicocca: Universita degli Studi di Milano-Bicocca, Milan, Italy

features (PPF) for feature extraction and analysis. Lastly, 3DSmoothNet is a very efficient descriptor built using a voxelized smoothed density value representation with a Siamese deep learning architecture and fully convolutional layers [13].

While through the years many different kinds of feature descriptors for global point clouds registration have been proposed, they all share a common problem. Since they aim at representing the geometric structure of an environment, they do not work properly in the absence of such structure. Such situation is quite common in many natural environments, for which the most common feature-based solutions do not work properly, as highlighted by previous works [14].

PHASER is a global point cloud registration algorithm which, contrary to most state-of-the-art approaches, is not based on features and which does not rely on finding correspondences between the points of the source and the target point clouds [15]. Instead, it exploits the Fourier transform of the point clouds to calculate a rigid transformation that maximizes the cross-correlation, a measure of similarity, between the two clouds. In addition, PHASER is able to use data other than range, such as the intensity channel of a LiDAR, to improve the registration result. This correspondence-free approach does not rely on local features, which can be greatly affected by noise, distortion, or outliers. Therefore, PHASER is particularly suitable for difficult registration problems, such as those dealing with noisy point clouds, low overlap, and different density between the two clouds. It also performs very well on challenging datasets representing natural environments.

Despite its good performances on unstructured point clouds depicting natural environments, PHASER shares most of the drawbacks of most global point clouds registration techniques which make it works not always work optimally:

- The objective function in global point clouds registration is often not exact, but derived from some kind of heuristics. This means that, even when the global optimum is found, there is no guarantee that it actually corresponds to the best alignment. On the contrary, it can happen that the best alignment corresponds to a local maximum smaller than the global one;
- Real-world point cloud data often contain noise and outliers. Moreover, the overlap between the two point clouds to align is often partial. These characteristics make the previous problem even worse.

This last point is particularly pertinent for PHASER, which, despite generating multiple solution hypotheses, ultimately relies on a single global optimum. However, as demonstrated, this global optimum may not always yield the best alignment.

To address this issue, we propose an extension to PHASER that considers multiple solution hypotheses, accompanied by a criterion to identify the most plausible among them. Our extension surpasses both PHASER and a state-of-the-art method (FPFH features combined with TEASER++) on challenging datasets like DARPA SubT and Newer College. These datasets, selected for their difficulty and similarity to those used in testing the original PHASER, incorporate additional data channels like intensity or reflectivity, besides point positions. We demonstrate the effectiveness of our proposal, particularly in scenarios where point clouds are very misaligned, representing the most challenging registration tasks. The most important improvements in our work include:

- We show that the global optimum used by PHASER does not necessarily correspond to the best alignment. By using a solution that corresponds to a local optimum rather than the global optimum, we can effectively achieve a better alignment. This is especially the case for the most difficult problems.
- Unlike PHASER, which relies solely on a single global optimum, the extension mitigates this limitation by exploring multiple plausible solutions. This addresses the inherent challenges in point cloud registration, especially in the presence of noise, outliers, or incomplete overlaps.
- We develop a criterion to identify the most plausible solution among multiple hypotheses, that is, a criterion to choose the best solution among multiple local optima. This ensures that the final alignment is not only globally optimal but also reflects the most suitable transformation for the given point cloud data, leading to more precise registration results.
- The extension demonstrates superior performance, particularly on challenging problems in the DARPA SubT and Newer College datasets. These, known for their complexity and variability, showcase the robustness and effectiveness of the proposed approach in real-world scenarios, surpassing both PHASER and other state-of-the-art methods.
- With its improved robustness and accuracy, the extension extends the applicability of PHASER to a wider range of environments and scenarios. Whether dealing with challenging natural landscapes, urban environments, or industrial settings, the proposed approach offers enhanced performance and adaptability across different domains.

Overall, the proposed extension represents a significant advancement over the original PHASER algorithm, offering a more robust, accurate, and versatile solution for global point cloud registration tasks in complex and dynamic environments (Fig. 1).

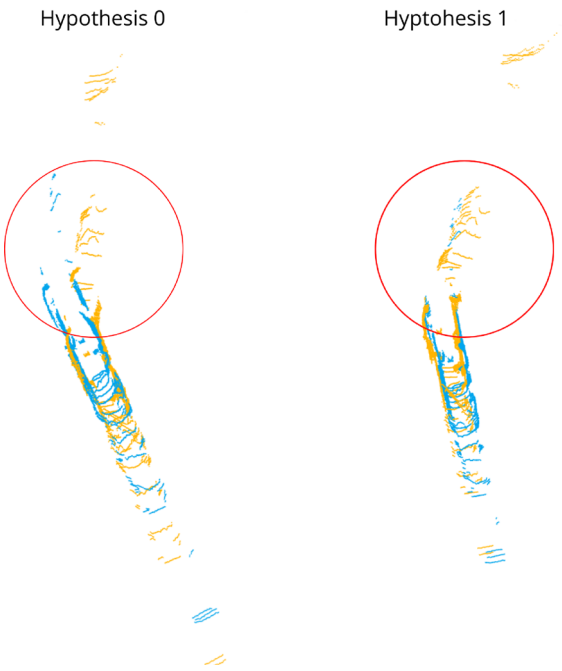

Fig. 1 A registration problem from the DARPA SubT dataset, where the second hypothesis is better than the first (original) result

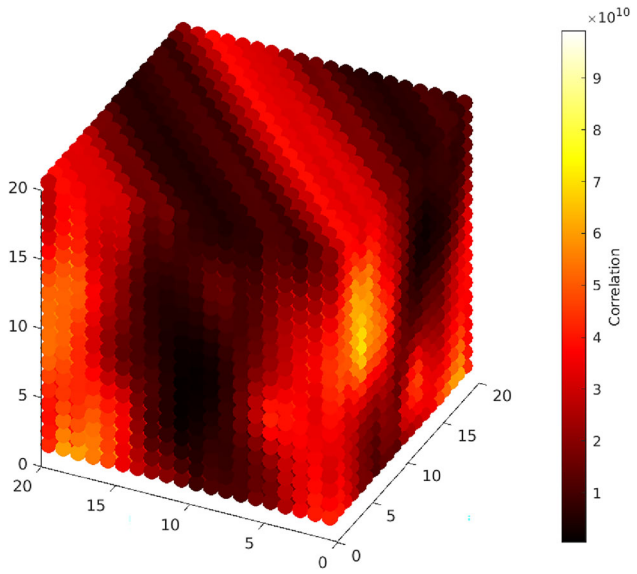

Fig. 2 Graphical representation of a correlation tensor. Each element corresponds to a combination of 3 rotation angles or 3 translation vectors. The color of the elements represents the correlation between the source point cloud, after applying the transformation, and the target point cloud

2 Methods

Since our proposal is an extension of PHASER, we briefly introduce its working. For the details, please consult the original work [15].

PHASER analyzes point clouds in the frequency domain and calculates rotation and translation correlation tensors to extract optimal rotation and translation values. The rotation correlation is computed by projecting the two point clouds onto a sphere, obtaining a $N \times N \times N$ tensor, where N depends on the bandwidth used for the spherical Fourier transform. Along each dimension of the tensor, an index represents a rotation angle. Since the rotation tensor is 3-dimensional, each element represents a rotation expressed by the convention of ZY'Z"-Euler angles. The value of an element of the tensor is the correlation between the source and target point clouds when the corresponding rotation is applied to the source point cloud. To compute the translation correlation, PHASER uses the two point clouds, re-sampled with fixed grid, and produces an $M \times M \times M$ tensor, with M depending on the bandwidth used for the spatial Fourier transform, where the indexes of the tensor correspond to translations along the x, y, and z axes. The value of an element of the tensor is the correlation between the point clouds when the corresponding translation is applied to the source point cloud.

PHASER uses these correlation tensors to estimate the rigid transformation that best aligns two point clouds. First, the rotation correlation tensor is computed and the maximum is extracted to find the best rotation. Then, this rotation is applied to the source point cloud; the translation correlation tensor is computed and used to find the translation that maximizes the correlation.

2.1 Multiple hypothesis generation

The central idea of this work is that the correlation tensors have multiple local maxima representing multiple rotation and translation hypotheses.

The tensors are structured as follows: each element in the tensor corresponds to either a rotation or a translation, and the value of each element represents the correlation value between the source point cloud, after applying the transformation, and the target point cloud, as shown in Fig. 2. It is easy to see that the spatial arrangement of the elements of the tensors is of fundamental importance in finding a meaningful structure of the correlation trends.

We leverage the structured of tensors to identify local maxima in both the rotation and translation tensors. Using GPU parallelization with libtorch [16] tensors and CUDA, we implement an efficient approach to extract peaks. This involves a max pooling operation, where the kernel size defines the neighborhood around a peak. Essentially, a location in the tensor is a peak if its value exceeds that of all neighboring elements. Subsequently, we generate a mask by comparing the max-pooled tensor with the original, identifying the local maxima. To speed-up the following steps, we discard peaks with values below half the global maximum.

To speed-up the computation of the subsequent steps, we discard all the peaks which have a value smaller than half of the global maximum. The total execution time of our proposal includes the following components:

- the time for the computation of the spherical Fourier transform of the target point cloud ($T_{\text{FFT\_sph\_target}}$);
- the time for the computation of the spherical Fourier transform of the source point cloud ($T_{\text{FFT\_sph\_source}}$);
- the time for the original PHASER rotation cross-correlation calculation (T_{rot});
- the time for extracting rotation peaks ($T_{\text{rot\_peak\_extraction}}$);
- the time for the computation of the spatial Fourier transform of the target point cloud ($T_{\text{FFT\_spat\_target}}$);
- the time for the computation of the spatial Fourier transform of the source point cloud, calculated for each rotation peak ($T_{\text{FFT\_spat\_source}}$);
- the time for the original PHASER translation cross-correlation calculation, calculated for each rotation peak (T_{trans});
- the time for extracting translation peaks ($T_{\text{trans\_peak\_extraction}}$).

Given these components, the total execution time T_{total} is:

$$\begin{aligned} T_{\text{total}} = {} & T_{\text{FFT\_sph\_target}} + T_{\text{FFT\_sph\_source}} \\ & + T_{\text{rot}} + T_{\text{rot\_peak\_extraction}} + T_{\text{FFT\_spat\_target}} \\ & + (T_{\text{FFT\_spat\_source}} \times N_{\text{rot\_peaks}}) + (T_{\text{trans}} \times N_{\text{rot\_peaks}}) \\ & + T_{\text{trans\_peak\_extraction}} \end{aligned}$$

where $N_{\text{rot\_peaks}}$ is the number of rotation peaks.

The time requirements for the spherical and spatial Fourier transform, and the computation of the cross-correlation tensors are the same as the original PHASER. The computational time of the peak search, according to the PHASER's parameters used in the experiment section and executed on a NVidia GTX 1080ti GPU, is equal to 0.3839 s for the rotation tensor, and to 0.3851 s for all the 4 translation tensors concurrently.

Our extended algorithm first computes the rotation correlation tensor. Then, for each peak, the corresponding rotation is applied to the source point cloud. For each rotated point cloud, a translation correlation tensor is computed, multiple peaks are extracted, and the corresponding translations are applied to the source point cloud to obtain the final transformation hypothesis, as shown in Fig. 3.

2.2 Local refinement and hypothesis choice

According to the above methodology, the transformations considered by the technique are basically sampled from a discrete grid. Therefore, the result might not be very accurate. Moreover, applying a local refinement step after a global registration is a commonly used technique to achieve a finer registration. For these reasons, we refine each hypothesis with the well-known local registration algorithm ICP.

We used the libpointmatcher ICP implementation, which allows, in addition to the standard algorithm, an outlier filtering step [17]. Outlier filtering allows to reject wrong correspondences found by the closest point policy, by associating a weight to each correspondence found. The weight is then used to weight the contribution of the correspondence to the function to optimize. According to the work of Babin et al. [17], we chose the Cauchy MAD function as the outlier filter.

After applying ICP, we compute the weighted residual error between the source and target point clouds. In particular, we use the point-to-point distance between correspondences weighted by the outlier rejection function, not just the point-to-point distance, that is,

$$\text{err} = \sum_{i=1}^{N} w_i \| s_i - t_i \|^2 \tag{1}$$

where N is the total number of correspondences, s_i is a point from the source point cloud, t_i is the closest point in the target point cloud, and w_i is the weight computed using the outlier rejection function.

After the local refinement is applied to each hypothesis and the residual is computed, we take the hypothesis which has the least residual error as final result of our extended algorithm.

3 Experimental setup

Since the original version of PHASER has been evaluated using data which is not public, we could not reproduce the exact same experiments. Therefore, we used two different datasets for our experiments: the "Newer College" dataset [18], and the "CERBERUS DARPA SubT Challenge" dataset [19, 20]. The former was acquired by Ramenzani et al. using an Ouster OS1 LiDAR (64 channels), which is the same LiDAR used for the original experiments. The data include the coordinates of the points, along with intensity, reflectivity and "ambient" data (the last two are specific to Ouster LiDARs), and capture an outdoor environment in a university campus. The second dataset was acquired by the CERBERUS team during the DARPA SubT Challenge using a Velodyne VLP-16 Puck LITE, which is a lower resolution LiDAR compared to the Ouster, and was obtained in underground tunnels and structures. These datasets allowed us to evaluate our approach both in a setup similar to the original experiments and in a more challenging one.

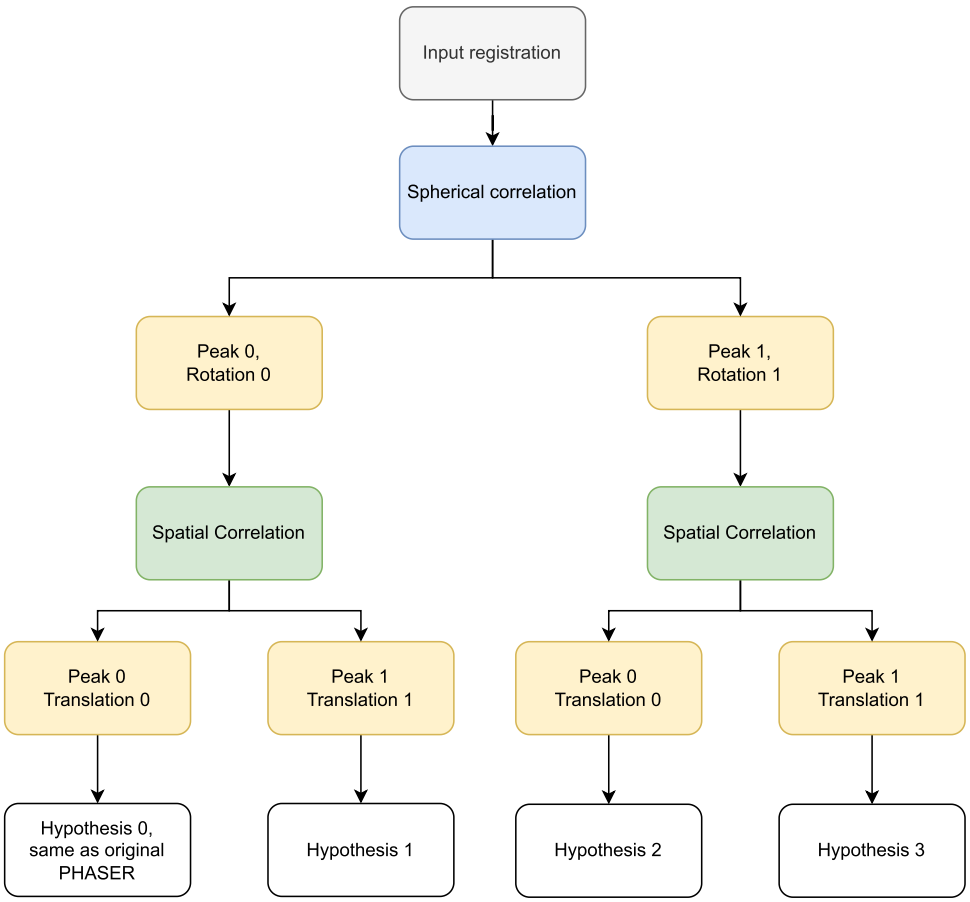

Fig. 3 Hypotheses generation pipeline. In this example, only the first 2 largest rotation peaks and the first 2 largest translation peaks are taken into account

For the evaluation, a set of experiments for each dataset was generated, similar to the "real-world" experiments of the original work. A fundamental characteristic of a registration problem is the distance between the poses of the sensors at the acquisition of the two point clouds. This is because farther poses correspond to a larger misalignment and, therefore, to a more challenging problem. For this reason, we want to measure how the performance of an approach varies when the problem becomes more challenging. Therefore, using the two datasets, we generated 400 registration problems, corresponding to pairs of point clouds generated to uniformly sample the space of possible different poses between the source and target point clouds. For the "Newer College" dataset, the distance range used is from 0 to 5 ms, while for the "DARPA SubT Challenge" dataset, the range used is from 0 to 2.5 ms. These ranges were selected according to each dataset, to guarantee a sufficient amount of overlap between the point clouds to align. Since we are testing a global registration approach, no initial guess on the transformation to estimate was provided.

Using the same pair of point clouds, we also propose additional experiments aimed at evaluating the performance of our proposal with noisier or sparser (that is, less points per volume) data.

For the first type of experiments, we sample the noise from a Gaussian distribution and add it to the points of the point clouds. The Gaussian has a mean value of zero and a standard deviation calculated at runtime, to obtain a defined Peak Signal-To-Noise ratio (PSNR). In particular, similar to the original PHASER experiments [15], we used three different PSNR values: 40, 45, and 50db. The noise is added to all the point clouds channels, hence both on the coordinates and on additional channels, such as intensity.

To generate experiments with sparser point clouds, instead, we randomly downsample the target point cloud to remove 50% of the points.

To measure the quality of the registration, we use the metric defined in the Point Clouds Registration Benchmark [14]: given a source point cloud S, consisting of n points s_i, the rigid transformation T estimated by the registration algorithm, and the ground truth transformation G, the error is calculated by measuring the distance between the source point cloud aligned using the ground truth and the source point cloud aligned using the estimated transformation. The distance is defined as follows:

$$D(G \cdot S, T \cdot S) = \frac{\sum_i \frac{\|G \cdot s_i - T \cdot s_i\|}{\|s_i - \overline{S}\|}}{n}$$

Table 1 Estimated ground truth accuracy for the datasets used

Dataset	Mean error [m]	Std. Dev [m]
Newer College	0.14	0.05
Darpa SubT Challenge	0.03	0.02

where $\overline{S}$ is the centroid of S, $T \cdot S$ is the application of the transformation T to S, and $\|x\|$ is the L_2 norm of the vector x. D represents the average distance between the ground truth and the estimated positions of a point, normalized by the average distance to the centroid of the point cloud.

While different metrics to evaluate point clouds registration algorithms exist, this was chosen because it combines the rotation and translation errors so that different approaches can be compared. Otherwise, e.g., if one approach had a lower translation error and the other a lower rotation error, there would be no objective way to decide which approach gives the best result. For more details, please consult the corresponding paper [14].

Moreover, we evaluated the accuracy of the ground truth according to the method proposed by [14]. For each registration problem, we align the source and target clouds, in their ground truth pose, using the registration algorithm introduced by Agamennoni et al. [21]. The metric discussed above is used to measure the ground truth error. The results are presented in Table 1, and, according to this measure, we consider the ground truth of both datasets good enough to be used for our experiments.

4 Results

Since we propose an extension to the original PHASER approach, we used similar parameters for our experiments. Specifically, we set the spherical bandwidth parameter $\tilde{B} = 120$ and the spatial bandwidth parameter $B = 5$. We started from the same parameters used for the original testing activity and fine-tuned them to get the best results. Fine-tuning was necessary to obtain the best results with PHASER, because we use different datasets for our experimental activity. Moreover, we used range and intensity channels for the experiments on the DARPA dataset, and range, intensity, ambient, and reflectivity for the Newer College dataset. Concerning the multiple hypotheses generation of our proposal, we limit our search to the first 4 local maxima for both rotation and translation, therefore considering a total of 16 transformations.

We also compared both the original PHASER and our variant with a state-of-the-art feature-based approach. Considering their results on natural datasets [14], we used FPFH features [5] to look for correspondences, along with TEASER++ [22] to estimate a transformation.

4.1 Results without ICP refinement

Figures 4 and 5 show the results of our approach, compared to those obtained using FPFH features in conjunction with TEASER. The figures allow to compare how the performance of the approaches varies with respect to the initial misalignment between the point clouds (depicted by the black dashed line). On the x-axis, there is the distance between the source and target point cloud. This distance corresponds to the distance between the two positions of the sensor at the acquisition. On the x-axis, we have the error, as defined in the previous section. Of course, not every distance is represented in our experimentation (much like in the original). For this reason, for each distance on the x-axis, we show the result in terms of a moving average, calculated using a window with a size of 20 elements.

Our proposal generates a set of hypotheses rather than a single solution. In the full version of the approach, these hypotheses are refined with ICP and evaluated according to an ICP-like metric. With this first set of experiments, however, we want to show that there is often a better solution than the one generated by PHASER and that our approach is capable of generating it. This leads to the need for an extension of PHASER. For this reason, in this section we evaluate our proposal without an ICP refinement step, using the best hypotheses selected according to the ground truth as the outcome. This is of course not representative of the actual performance of our approach, which is shown in Figs. 6 and 7.

Considering Figs. 4 and 5, we can see that our proposal (red line) and the original PHASER implementation (blue line) perform very similarly when the distance between the two point clouds is below a threshold of about 3.0 meters for the Newer College dataset and of 0.5 meters for the Darpa dataset. Beyond such value, the error of the original PHASER increases severely, while that of our extension remains much lower. Therefore, we can state that while the two versions perform similarly with easier problems, when the distance increases, there are better solutions that are discarded by PHASER, while our extension is able to generate solutions that allow registration under much more challenging conditions. This is the case, for example, if the acquisition frequency is too low in relation to the speed of the observer and thus the distance between the point clouds to align increases.

Similar observations hold also when considering an approach based on FPFH features and TEASER (green line). This approach is very effective for easier problem, but does not allow accurate registration of distant point clouds.

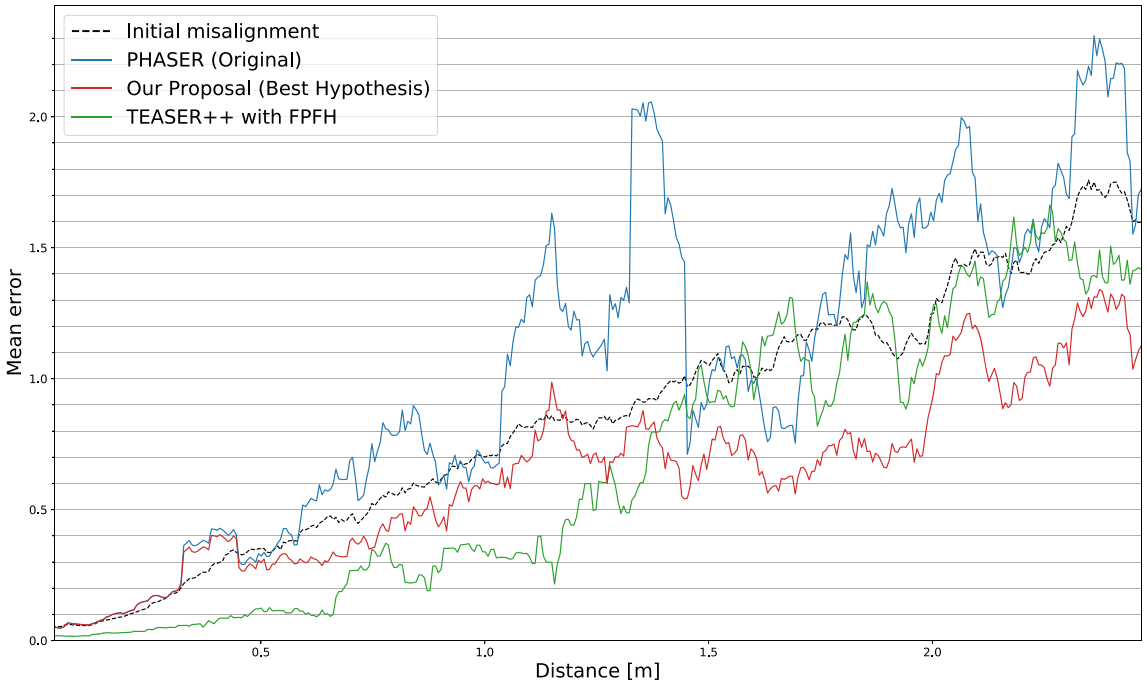

Fig. 4 Residual error, w.r.t. the initial distance between the point clouds, of three approaches we compared on the DARPA dataset: TEASER++ with FPFH features, the original PHASER implementation, and our extension. The dashed line represents the initial misalignment between the point clouds, measured with the same metric of the errors

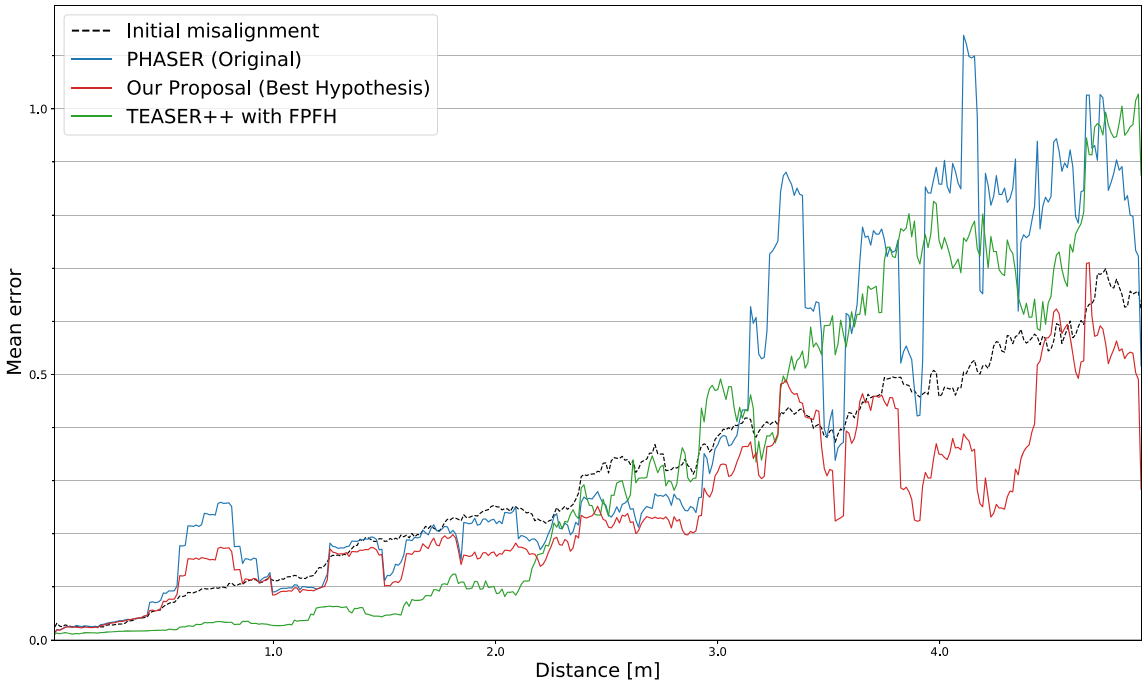

Fig. 5 Residual error, w.r.t. the initial distance between the point clouds, of three approaches we compared on the Newer College dataset: TEASER++ with FPFH features, the original PHASER implementation, and our extension. The dashed line represents the initial misalignment between the point clouds, measured with the same metric of the errors

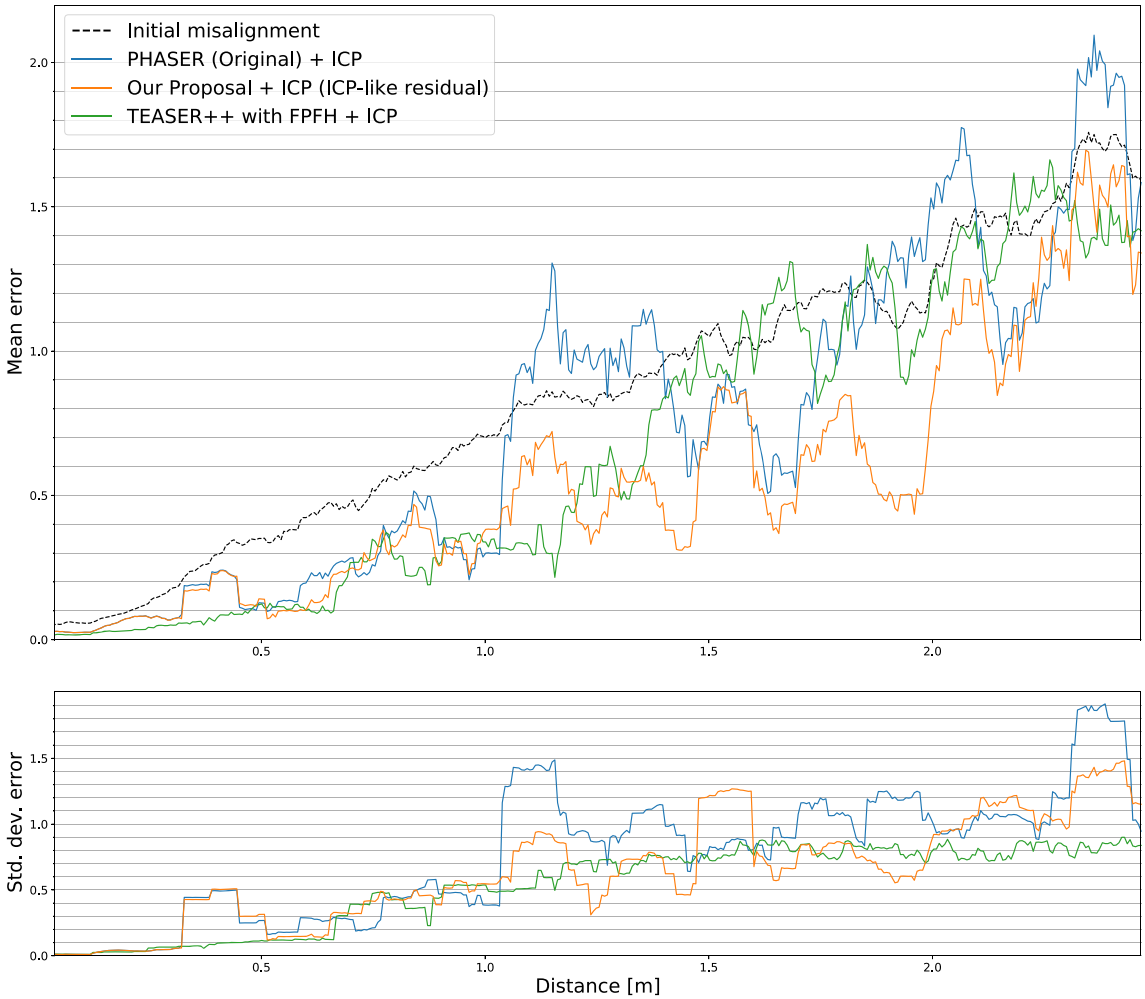

Fig. 6 Residual error, w.r.t. the initial distance between the point clouds, of three approaches we compared on the DARPA dataset: TEASER++ with FPFH features, the original PHASER implementation, and our extension. The dashed line represents the initial misalignment between the point clouds, measured with the same metric of the errors. In the top picture, we have the moving average of the errors, calculated with a window of 20 elements. Below, we have the corresponding standard deviation

4.2 Results with ICP refinement

Expanding the experiments presented in the previous sections, we also tested the aforementioned approaches in conjunction with a refinement step performed using ICP, which is a very common use case. When aligning point clouds with a very large initial misalignment, it is quite common to perform a first initial registration step to compute a rough estimate of the transformation. This first rough estimate is then followed by a refinement step, which should provide a much more accurate solution, since the search space becomes smaller. This refinement step is often performed using ICP or ICP-like approaches.

The results of this second set of experiments on both datasets are shown in Figs. 6 and 7.

Each plot shows the initial misalignment, the hypothesis chosen using the ICP residual, which is the actual outcome of our approach, and the result of TEASER++ and FPFH features. To allow for a fair comparison, the same ICP configuration was used to refine the results of TEASER++. Similar to the previous set of experiments, the three approaches perform similarly on simpler problems, i.e., below a certain distance. However, our proposal outperforms the other approaches when it comes to solving more challenging problems. It should be noted that the results on the two datasets are slightly different. On the Newer College dataset, the combination of high-resolution LiDAR and open spaces allows for good sampling and high-quality subsequent frequency analysis. On the DARPA dataset, on the other hand, the combination of a lower- resolution LiDAR and the tunnel-like

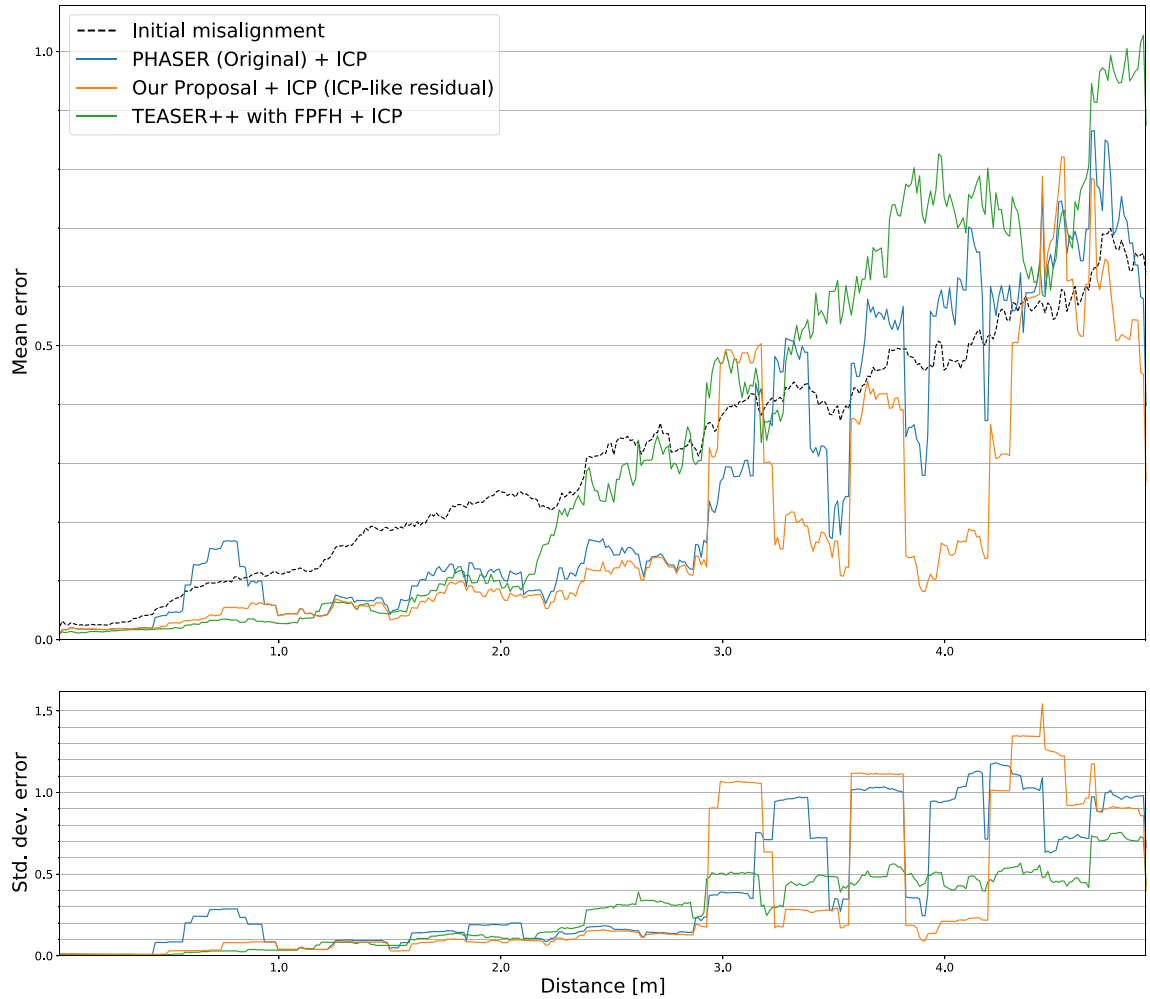

Fig. 7 Residual error, w.r.t. the initial distance between the point clouds, of three approaches we compared on the Newer College dataset: TEASER++ with FPFH features, the original PHASER implementation, and our extension. The dashed line represents the initial misalignment between the point clouds, measured with the same metric of the errors. In the top picture, we have the moving average of the errors, calculated with a window of 20 elements. Below, we have the corresponding standard deviation

environment might not always allow for sufficiently high frequencies required for accurate registration.

Besides the distance between the positions of the sensors, a critical parameter which defines the difficulty of a registration problem is the overlap between the point clouds. We determined the level of overlap by measuring the percentage of points within a point cloud that matched with corresponding points in another point cloud, which had been aligned using the ground truth [14]. Given that not all points had an exact counterpart, we considered two points to form a correspondence if their distance fell below a threshold of 0.1 meters. In Figs. 8 and 9, we show how the residual error of the three approaches varies according to the overlap between the source and target point clouds. In both datasets, there seems to be no relevant difference between the approaches in terms of mean residual error when the overlap is too low (less than 50%). Furthermore, if we look at the standard deviation bars in Figs. 8 and 9, we conclude that there is a huge variability in the quality of the results for each approach. The variability is so large that the differences in the mean errors are not significant.

4.3 Other experiments

The outcomes of our experiments where noise was introduced into the target point clouds are illustrated in Figs. 12 and 13. An analysis of these results confirms that our proposed method outperforms both the original PHASER and TEASER++ when combined with FPFH features. This trend remains consistent with experiments conducted without additional noise, with the performance gap between our approach and others widening as problem difficulty

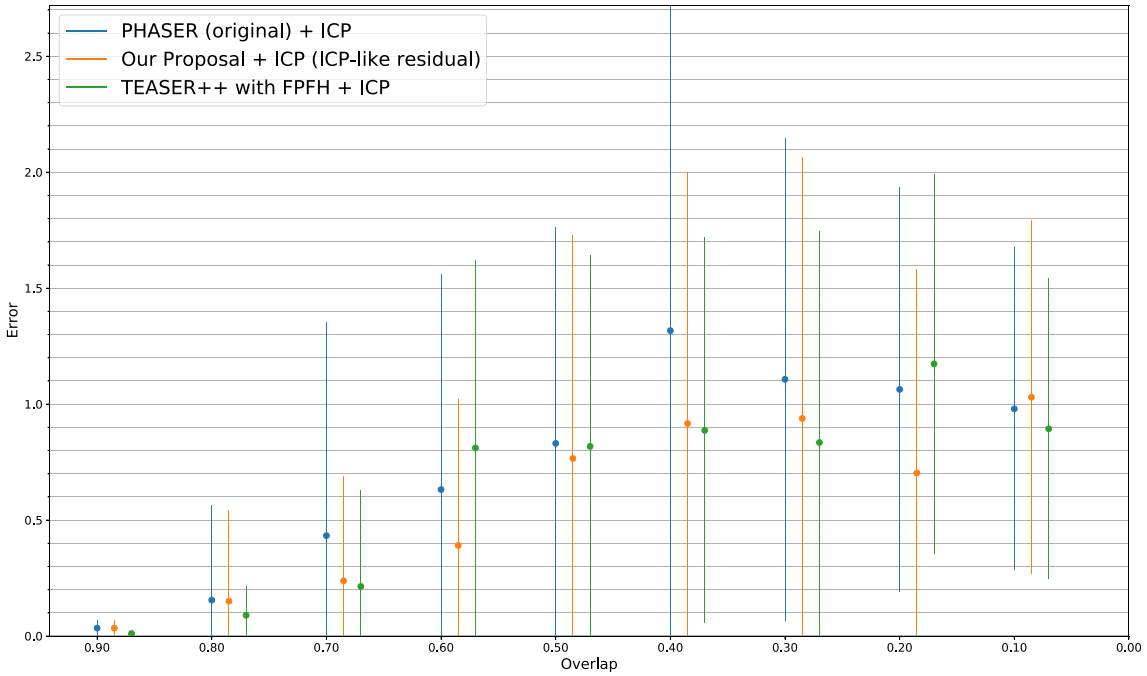

Fig. 8 Comparison of the error on the DARPA dataset w.r.t. the initial misalignment, dividing the experiments by the overlap between source and target pointcloud. The dots represent the average error, while the lines depict the standard deviation. The results with overlap less than 0.1 are not shown due to their low number

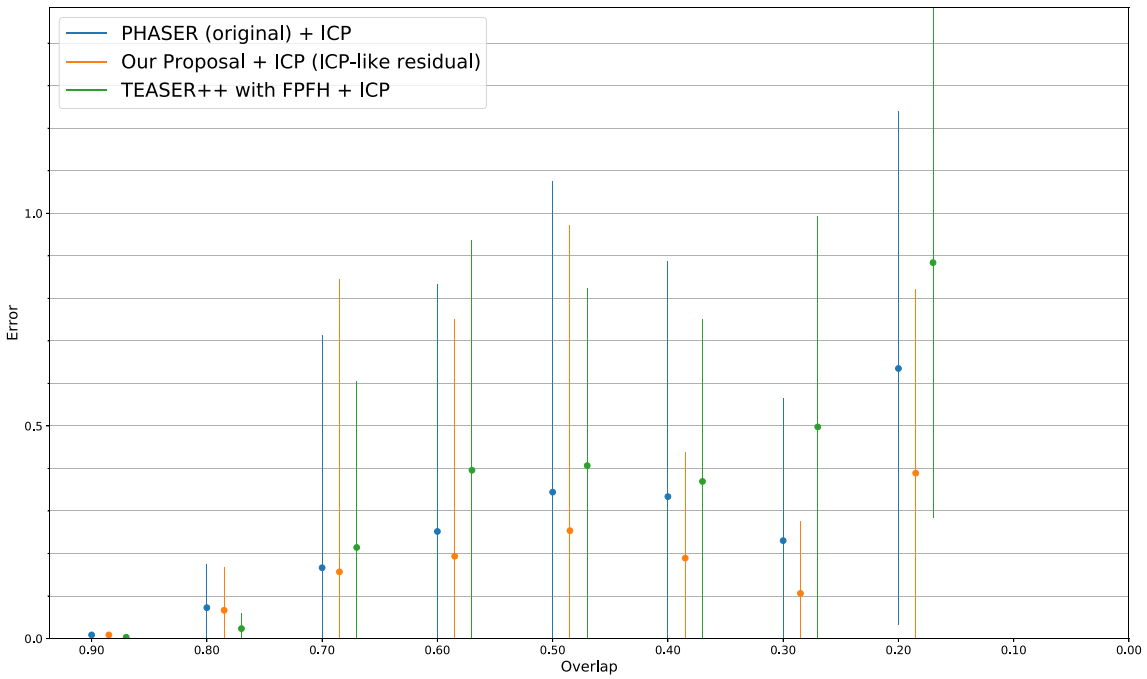

Fig. 9 Comparison of the error on the Newer College w.r.t. the initial misalignment, dividing the experiments by the overlap between source and target pointclouds. The dot represents the mean error, and the lines the standard deviation.The results with overlap less than 0.2 are not shown due to their low number

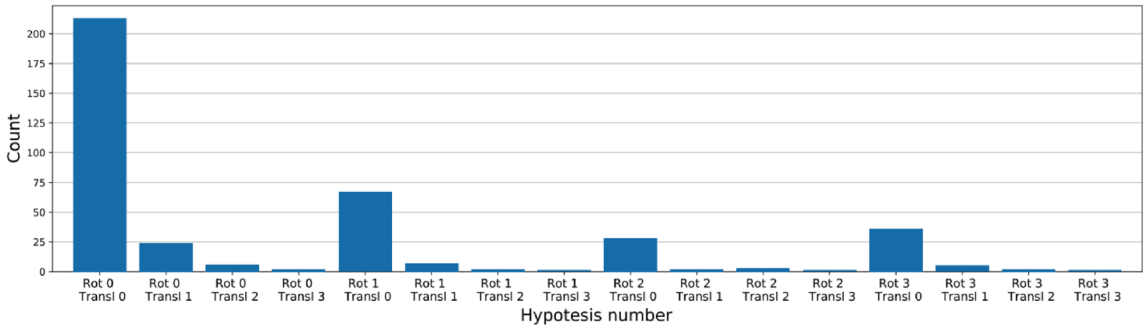

Fig. 10 Histogram illustrating the frequency of optimal rotation and translation combinations on the DARPA Challenge dataset. "Rot0 Transl0" denotes the combination of the highest rotation correlation tensor peak with the highest translation correlation tensor peak. "Rot0 Transl1" represents the combination of the highest rotation peak with the second highest translation peak, and so forth

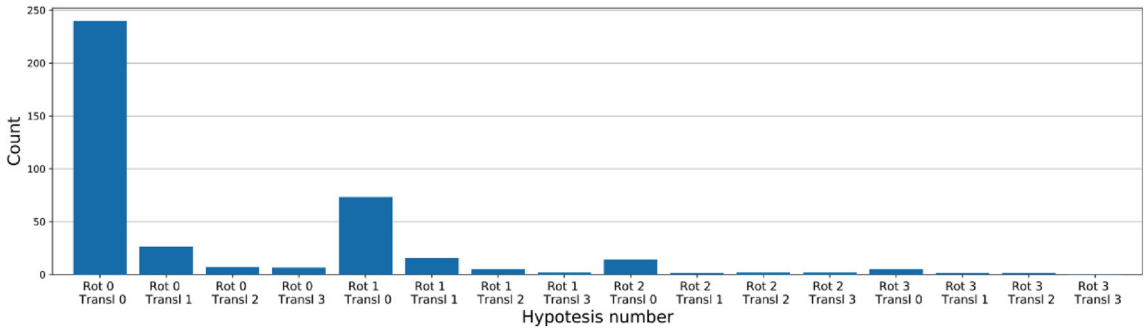

Fig. 11 Histogram illustrating the frequency of optimal rotation and translation combinations on the Newer Colleger dataset. "Rot0 Transl0" denotes the combination of the highest rotation correlation tensor peak with the highest translation correlation tensor peak. "Rot0 Transl1" represents the combination of the highest rotation peak with the second highest translation peak, and so forth

increases. Notably, TEASER++ struggles to yield satisfactory solutions across most scenarios, as evidenced by its green line on the plot frequently surpassing the dashed black line indicating the initial misalignment. This highlights that employing TEASER++ with FPFH features often worsen the problem rather than solving it. This is surprising, considering TEASER's prior excellent results on challenging datasets [14], but underscores the ineffectiveness of feature-based strategies in environments lacking clear structural cues, such as underground tunnels.

As explained earlier, the rotation and translation correlation tensors have multiple local maxima. Differently from the original PHASER, we consider four maxima and choose the best one according to an ICP-like method. Figures 10 and 11 illustrate the frequency of selecting specific combinations of maxima for the Newer College and DARPA datasets. For instance, "Rot0 Transl0" denotes the combination of the highest rotation correlation tensor peak with the highest translation correlation tensor peak. "Rot0 Transl1" represents the combination of the highest rotation peak with the second highest translation peak, and so forth. As expected, most of the time the best solution was the combination of the rotation and translation global maxima. However, the advantage of considering multiple maxima cannot be neglected, as also demonstrated by the comparison of our approach with PHASER. This is particularly notable for the rotation correlation tensor, where the best rotation may correspond to the second or even third local maximum. In contrast, the translation benefits less from this strategy. Notably, considering the fourth local maximum does not yield significant advantages, leading us to limit our consideration to no more than four maxima for both the translation and rotation.

The results of the experiments, conducted after randomly sampling 50% of the points from the target point clouds, are shown in Figs. 15 and 14. These experiments highlight the challenge of aligning two point clouds with different densities, where the target point cloud is significantly sparser than the source. Even in such challenging scenarios, our proposed approach outperforms the original PHASER method. This aligns with our previous observations: the performance difference is less pronounced in easier problems but becomes significant in more difficult cases. Moreover, it is noteworthy that our proposal consistently improves the initial solution, even in very challenging problems (the orange line representing our proposal's result is always below the dashed black line representing the initial misalignment). In contrast, the

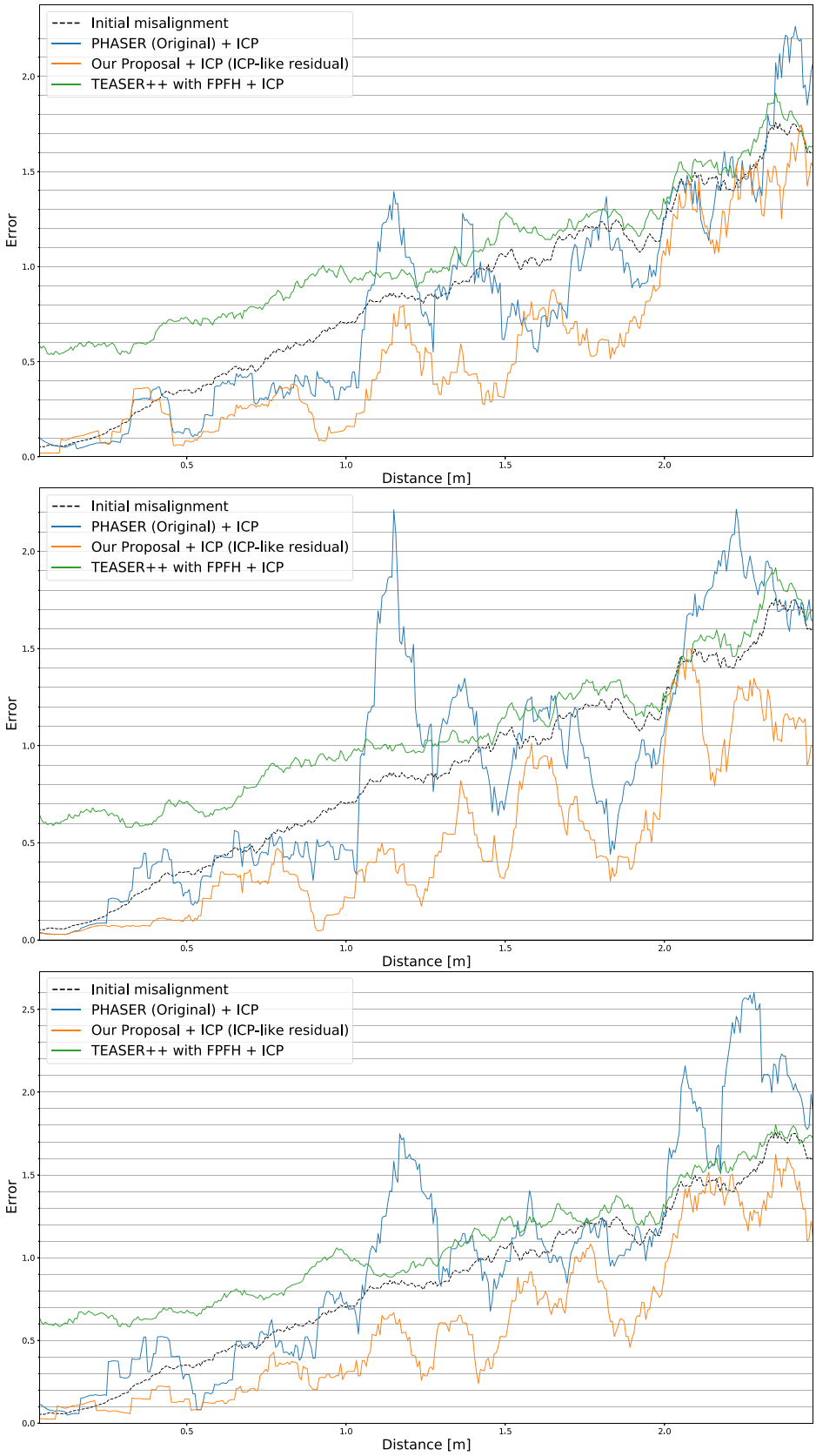

Fig. 12 Experiments on the DARPA dataset with different levels of added noise. From top to bottom PSNR 50, 45, and 40 dB

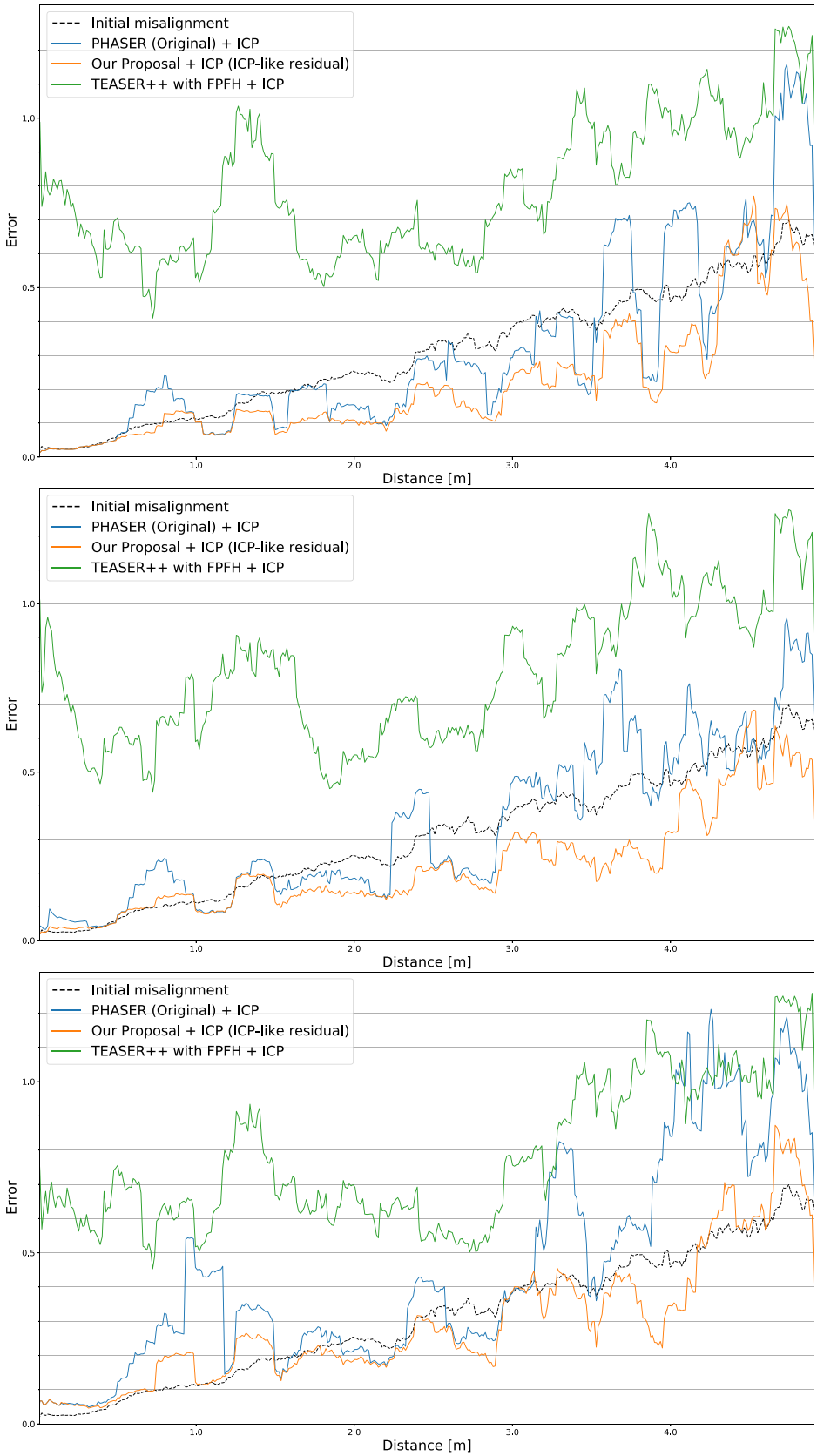

Fig. 13 Experiments on the Newer College dataset with added noise. From top to bottom PSNR 50, 45, and 40 dB

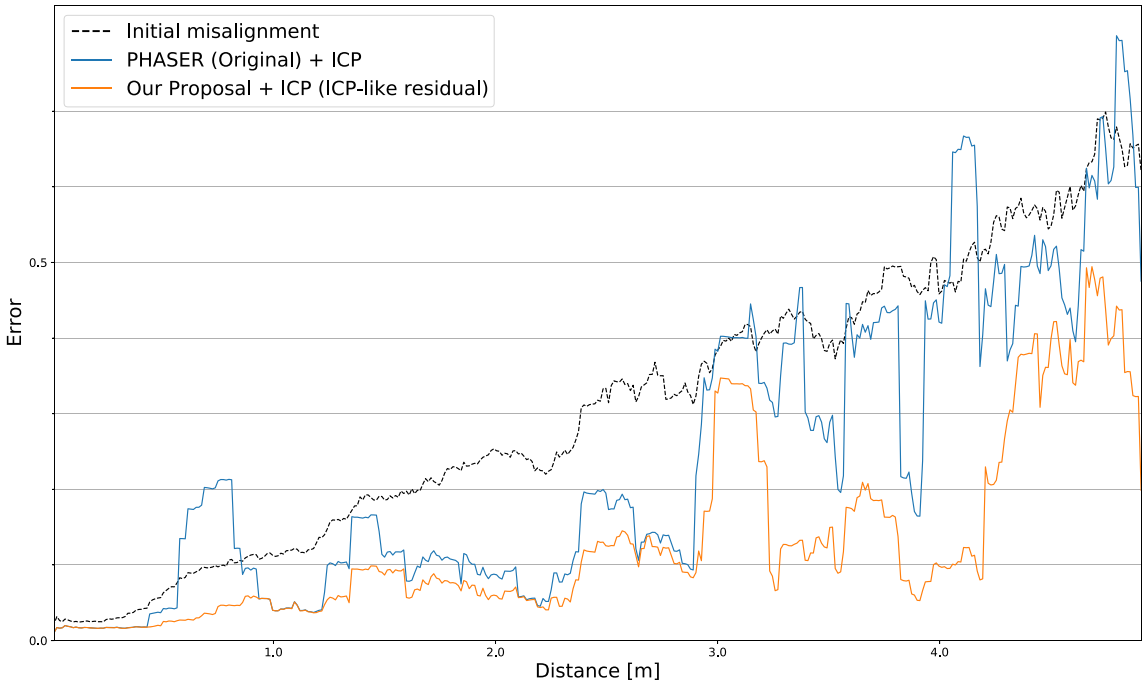

Fig. 14 Comparison of the proposed approach and the original PHASER on the Newer College Dataset, with the target point cloud randomly subsampled by 50%

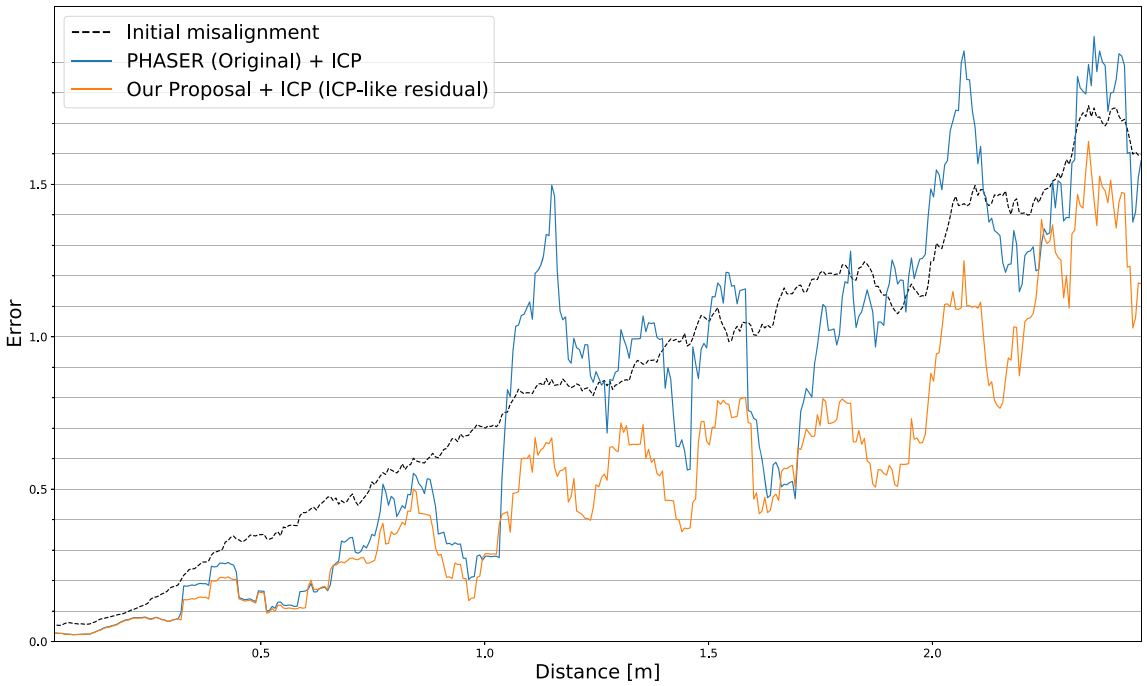

Fig. 15 Comparison of the proposed approach and the original PHASER on the DARPA dataset, with the target point cloud randomly subsampled by 50%

results of the original PHASER are much less consistent and not always better than the initial misalignment. We do not show the results of TEASER++ with FPFH features, as in the previous experiments, because using this type of feature with such a sparse target point cloud often led to a failure in the computation or produces features that are not sufficiently descriptive.

5 Conclusions

Our work highlights the open issue of global point cloud registration in challenging environments. This problem is of great importance due to its critical implications. To tackle this issue, we employed the PHASER registration algorithm, which avoids reliance on features and instead relies on cross-correlation, a choice driven by the limitations of feature-based methods in unstructured environments.

However, our investigation revealed that PHASER's performance falls short in certain scenarios, mainly due to cross-correlation's maxima that do not correspond with the optimal rigid transformation. To address this limitation, we extended the algorithm to evaluate multiple local maxima and select the most suitable one, employing a classical ICP-like scoring approach. Our experimental results, conducted on challenging real-world datasets, consistently demonstrate the improved performance of this extended version, particularly in more complex scenarios. Moreover, our proposal also outperforms an approach based on TEASER++ and FPFH features when solving more difficult problems consisting of distant point clouds. This is noteworthy as TEASER++ has shown excellent performance in a previous comparison [14]. This advancement in point cloud registration offers promise for critical applications in challenging environments.

Acknowledgements Editorial Policies for: Springer journals and proceedings: https://www.springer.com/gp/editorial-policies Nature Portfolio journals: https://www.nature.com/nature-research/editorial-policies *Scientific Reports*: https://www.nature.com/srep/journal-policies/editorial-policies BMC journals: https://www.biomedcentral.com/getpublished/editorial-policies.

Funding Open access funding provided by Universitá degli Studi di Milano - Bicocca within the CRUI-CARE Agreement.

Data availability https://anonymous.4open.science/r/phaser_extended_public-CC51/ (BLINDED).

Code Availability https://anonymous.4open.science/r/phaser_extended_public-CC51/ (BLINDED).

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Chen Y, Medioni G (1992) Object modelling by registration of multiple range images. Image Vis Comput 10(3):145–155
2. Zhang Z (1994) Iterative point matching for registration of free-form curves and surfaces. Int J Comput Vis 13(2):119–152
3. Besl PJ, McKay ND (1992) Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures, vol 1611. Spie, pp 586–606
4. Rusu RB, Blodow N, Marton ZC, Beetz M (2008) Aligning point cloud views using persistent feature histograms. In: 2008 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 3384–3391
5. Rusu RB, Blodow N, Beetz M (2009) Fast point feature histograms (fpfh) for 3d registration. In: 2009 IEEE international conference on robotics and automation, pp 3212–3217. https://doi.org/10.1109/ROBOT.2009.5152473
6. Jiang J, Cheng J, Chen X (2009) Registration for 3-d point cloud using angular-invariant feature. Neurocomputing 72(16–18):3839–3844
7. Sehgal A, Cernea D, Makaveeva M (2010) Real-time scale invariant 3d range point cloud registration. In: International conference image analysis and recognition. Springer, pp 220–229
8. Choy C, Park J, Koltun V (2019) Fully convolutional geometric features. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 8958–8966
9. Zeng A, Song S, Nießner M, Fisher M, Xiao J, Funkhouser T (2017) 3dmatch: learning local geometric descriptors from rgb-d reconstructions. In: CVPR
10. Yew ZJ, Lee GH (2018) 3dfeat-net: weakly supervised local 3d features for point cloud registration. In: ECCV
11. Charles RQ, Su H, Kaichun M, Guibas LJ (2017) Pointnet: deep learning on point sets for 3d classification and segmentation. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 77–85. https://doi.org/10.1109/CVPR.2017.16
12. Deng H, Birdal T, Ilic S (2018) Ppfnet: global context aware local features for robust 3d point matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 195–205
13. Gojcic Z, Zhou C, Wegner JD, Andreas W (2019) The perfect match: 3d point cloud matching with smoothed densities. In: International conference on computer vision and pattern recognition (CVPR)
14. Fontana S, Cattaneo D, Ballardini AL, Vaghi M, Sorrenti DG (2021) A benchmark for point clouds registration algorithms. Robot Auton Syst 140:103734. https://doi.org/10.1016/j.robot.2021.103734
15. Bernreiter L, Ott L, Nieto J, Siegwart R, Cadena C (2021) Phaser: a robust and correspondence-free global pointcloud registration. IEEE Robot Autom Lett 6(2):855–862. https://doi.org/10.1109/LRA.2021.3052418
16. Ansel J, Yang E, He H, Gimelshein N, Jain A, Voznesensky M, Bao B, Bell P, Berard D, Burovski E, Chauhan G, Chourdia A, Constable

W, Desmaison A, DeVito Z, Ellison E, Feng W, Gong J, Gschwind M, Hirsh B, Huang S, Kalambarkar K, Kirsch L, Lazos M, Lezcano M, Liang Y, Liang J, Lu Y, Luk C, Maher B, Pan Y, Puhrsch C, Reso M, Saroufim M, Siraichi MY, Suk H, Suo M, Tillet P, Wang E, Wang X, Wen W, Zhang S, Zhao X, Zhou K, Zou R, Mathews A, Chanan G, Wu P, Chintala S (2024) PyTorch 2: faster machine learning through dynamic python bytecode transformation and graph compilation. In: 29th ACM international conference on architectural support for programming languages and operating systems, vol 2 (ASPLOS '24). ACM. https://doi.org/10.1145/3620665.3640366. https://pytorch.org/assets/pytorch2-2.pdf

17. Babin P, Giguère P, Pomerleau F (2019) Analysis of robust functions for registration algorithms. In: 2019 International conference on robotics and automation (ICRA), pp 1451–1457. https://doi.org/10.1109/ICRA.2019.8793791
18. Ramezani M, Wang Y, Camurri M, Wisth D, Mattamala M, Fallon M (2020) The newer college dataset: Handheld lidar, inertial and vision with ground truth. In: 2020 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 4353–4360. https://doi.org/10.1109/IROS45743.2020.9340849
19. Tranzatto M, Dharmadhikari M, Bernreiter L, Camurri M, Khattak S, Mascarich F, Pfreundschuh P, Wisth D, Zimmermann S, Kulkarni M, Reijgwart V, Casseau B, Homberger T, De Petris P, Ott L, Tubby W, Waibel G, Nguyen H, Cadena C, Buchanan R, Wellhausen L, Khedekar N, Andersson O, Zhang L, Miki T, Dang T, Mattamala M, Montenegro M, Meyer K, Wu X, Briod A, Mueller M, Fallon M, Siegwart R, Hutter M, Alexis K (2022) Team CERBERUS wins the DARPA subterranean challenge: technical overview and lessons learned. arXiv. https://doi.org/10.48550/ARXIV.2207.04914. arXiv:2207.04914
20. Tranzatto M, Miki T, Dharmadhikari M, Bernreiter L, Kulkarni M, Mascarich F, Andersson O, Khattak S, Hutter M, Siegwart R, Alexis K (2022) Cerberus in the darpa subterranean challenge. Sci Robot 7(66):9742. https://doi.org/10.1126/scirobotics.abp9742
21. Agamennoni G, Fontana S, Siegwart RY, Sorrenti DG (2016) Point clouds registration with probabilistic data association. In: 2016 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 4092–4098. https://doi.org/10.1109/IROS.2016.7759602
22. Yang H, Shi J, Carlone L (2021) Teaser: fast and certifiable point cloud registration. IEEE Trans Rob 37(2):314–333. https://doi.org/10.1109/TRO.2020.3033695