https://doi.org/10.1016/j.isprsjprs.2020.11.014

Robust Global Registration of Point Clouds by Closed-Form Solution in the Frequency Domain

Rong Huang[a], Yusheng Xu[a,*], Wei Yao[b], Ludwig Hoegner [a], Uwe Stilla[a]

[a]*Photogrammetry and Remote Sensing, Technical University of Munich (TUM), Munich, Germany*

[b]*Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

Abstract

Point cloud registration is invariably an essential and challenging task in the fields of photogrammetry and computer vision to align multiple point clouds to a united reference frame. In this paper, we propose a novel global registration method using a robust phase correlation method for registration of low-overlapping point clouds, which is less sensitive to noise and outliers than feature-based registration methods. The proposed point cloud registration is achieved by converting the estimation of rotation, scaling, and translation in the spatial domain to a problem of correlating low-frequency components in the frequency domain. Specifically, it consists of three core steps: transformation from the spatial domain to the frequency domain, decoupling of rotation, scaling, and translation, and adapted phase correlation for robust shift estimation. In the first step, unstructured and unordered 3D points are transformed from the spatial domain to the frequency domain via 3D Fourier transformation, following a voxelization and binarization process. In the second step, rotation, scaling, and translation are decoupled by sequential operations, including Fourier transform, resampling strategies, and Fourier-Mellin transform. In the third step, the estimation of transformation parameters is transformed into shift estimation tasks. The shift estimation task is solved by a robust phase correlation method, in which low-frequency components are matched by decomposing the normalized cross-power spectrum and linearly fitting the decomposed signals with a closed-form solution by a $\ell 1$-norm-based robust estimator. Experiments were conducted using three different datasets of urban and natural scenarios. Results demonstrate the efficiency of the proposed method, with the majority of rotation and translation errors reaching less than 0.2 degrees and 0.5 m, respectively. Additionally, it is also validated by experiments that the proposed method is robust to noise and versatile to datasets with wide ranges of overlaps and various geometric characteristics.

Preprint submitted to ISPRS Journal of Photogrammetry and Remote Sensing November 15, 2020

Keywords: Point cloud registration, Fourier transforms, multidimensional phase correlation, low-frequency components, robust estimation

1. Introduction

Since the last decade, point clouds acquired via Light Detection and Ranging (LiDAR) or photogrammetric acquisition have frequently been used in a wide range of research fields and engineering projects (Vosselman and Maas, 2010). Point clouds were proposed to be the most proper data for 3D visualization in broader urban scenarios, owing to their capability of providing spatial coordinates of observed object surfaces, which disentangles tasks like interpretation and reconstruction of 3D scenes (Yang et al., 2013a; Huang et al., 2020c). However, in the observations using the laser scanning, in the scanned scene, only points in the path of laser beams can be measured, and points in the occluded or invisible area stay unconsidered. To overcome this drawback, we usually need to conduct multiple scans or potogrmmetric acquisitions to cover a large urban scene, mainly because of the occlusion by objects in urban scenes, such as, cars and buildings, and restricted locations of the scanners (Yang et al., 2016; Dong et al., 2018). Consequently, before any further use, a co-registration of these individually scanned point clouds becomes a vital task, ensuring full coverage of the entire scene (Dong et al., 2020).

Point cloud registration has long been a challenging work in the field of photogrammetry and computer vision, whose objective is to estimate a rigid transformation that aligns multiple individual but related point clouds into a unified coordinate system (Xu et al., 2019). These point clouds might be acquired from different viewpoints, at different times, using different platforms, or via multimodal sensors. Effective results of point cloud registration is the prerequisite of many applications, such as autonomous driving, 3D reconstruction (Lafarge and Mallet, 2012; Yang et al., 2013a), forest investigation (Polewski et al., 2019), construction monitoring (Bosché et al., 2015; Tuttas et al., 2017; Huang et al., 2020b), urban planning (Vosselman and Maas, 2010) and change detection (Gehrung et al., 2017; Hebel and Stilla, 2011; Hebel et al., 2013).

Generally, point cloud registration is achieved by identifying correspondences between point clouds, which is usually realized by a two-step solution. First, feature representation shall be established from original point clouds as the basis for the search of correspondences. Then, corresponding feature pairs can be identified based on the extracted features (Habib et al., 2010). Technically, on condition that correspondence between features are identified, the transformation parameters between the coordinate systems can be well estimated via optimization-based algorithms.

However, when operating point cloud registration following the aforementioned steps, several critical problems appear:

- Noise and outliers caused by temporary or moving objects: noise and outliers influence the dependability of some feature descriptors based on details of point clouds and even lead to failure in finding correspondence in some cases with low-quality point clouds;
- Uneven densities resulting from different viewing distances of scanners: for a terrestrial laser scanner (TLS), densities of point clouds decrease with the increase of observation distances, making the extracted features ineffectual;
- Incomplete data caused by occlusions in complex urban environments: the incompetence will result in the change of details of point clouds and thus lead to failure in search of corresponding feature pairs;
- Extensive data amount of point clouds: massive data size will lead to high computational effort and low time efficiency for most of point cloud registration algorithms;
- Self-similar and symmetric urban objects: the intricate and homogeneous architectural structures in urban scenarios could lead to the mismatching of correspondences, owing to similar and regular building and infrastructure elements;
- A low overlap ratio between point clouds: low overlaps will lead to defeats in finding sufficient feature pairs for estimating transformation parameters.

Considering these problems, we can find that the feature-based registration methods are greatly influenced by the quality of feature representations and the overlap ratio of point clouds.

To address the problems mentioned above, we aim to design a pairwise coarse registration framework that fulfills the following requirements: robust to noise and outliers and applicable to low-overlapping cases. In this paper, we provide an automatic and marker-free solution via a novel global registration method using a robust phase correlation method (GRPC). In the proposed method, point clouds are aligned with a transformation of seven degrees of freedom (DoFs) using global features generated in the frequency domain. Global features are deemed to be less easily influenced by low-overlapping issues and unevenly distributed point densities than features constructed based on local context (Huang et al., 2019, 2020a). Besides, high-frequency components which indicates noise and outliers in the 3D signals can be eliminated by representing 3D points using discrete signals and transforming them to the frequency domain. Compared to feature-based registration approaches (i.e., using key points or geometric primitives), global feature-based methods utilize underlying information provided by the whole point cloud, which provides adequate constraints for the geometric information (Xu et al., 2017) and simultaneously enhancing the reliability. Moreover, the ill-posed registration problem can be tackled by a straightforward estimation of phase angle differences, which provides a closed-form solution and simultaneously achieves a good balance between efficiency and effectiveness. In general, contributions and innovations of this work are listed as follows:

- We propose a new global feature-based point cloud registration framework, which is achieved by correlating low-frequency components of 3D signals presented by 3D point clouds, capable of dealing with high-frequency noise, low-overlapping cases, and small changes. Compared with local-correspondence-based strategy, the extraction of local features and correspondences is avoided, and the accuracy of registration does not rely on the intermediate step.

- We decompose the point cloud registration problem of seven DoFs into several multidimensional shift estimation sub-problems, which can be solved by a standard and closed-form solution, composing a sequence of Fourier-based transformations and optimizations.

- We propose a novel multidimensional shift estimation method based on subvoxel-based phase correlation, in which shifts are estimated by decomposing correlated cross-spectrum and fitting the decomposed signals using an l1 normalized line-fitting approach.

The remainder of this paper is organized as follows. In Section 2, a literature review on the mark-less point cloud registration methods is given. In Section 3, the principle of robust estimation of transformation between point clouds in the frequency domain is elaborated. Section 4 gives an application of the proposed principle, presenting a novel global point cloud registration method. Section 5 presents the experiments and evaluation, and Section 6 gives a detailed discussion and analysis of the obtained results. Finally, Section 7 concludes the paper and introduces the future work.

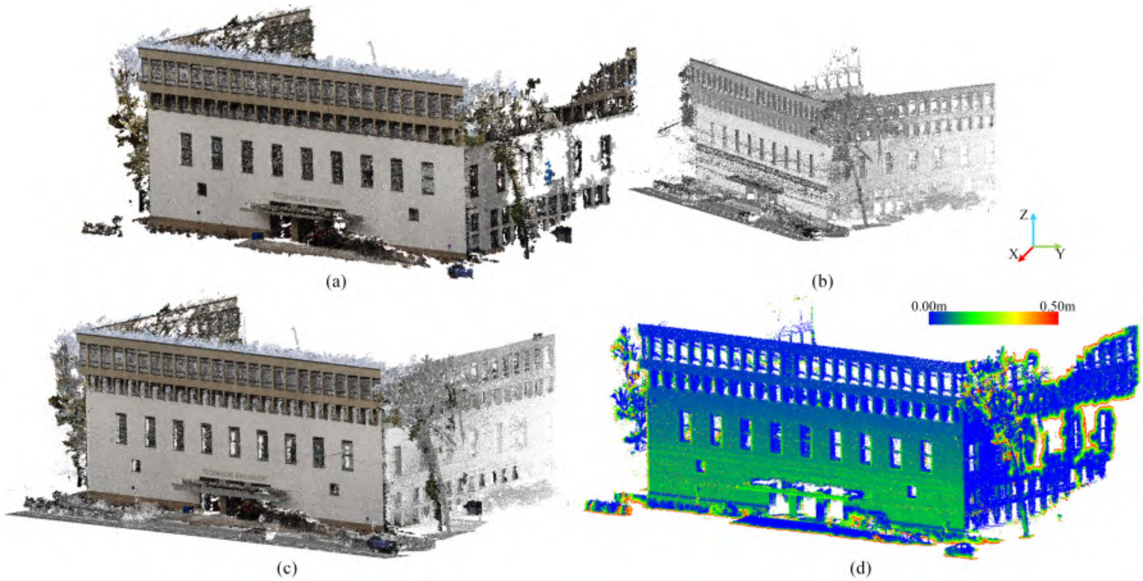

Figure 1: Illustration of registering multisource point clouds using our proposed method. (a) Photogrammetric and (b) laser scanning point clouds (from TUM-MLS-2016 dataset (Zhu et al., 2020)) with scaling changes and rotation. (c) Registered point clouds and (d) residual distances (rendered with colors) between corresponding points.

2. Related work

Numerous studies have been intensively reported to solve mark-less point cloud registration. Coarse registration and fine registration are the two major categories of registration approaches. In fine registration methods, iterative closest point (ICP) (Besl and McKay, 1992) and its variants, such as Geometric Primitive ICP (Bae and Lichti, 2008), geometric features + ICP (Gressin et al., 2013; Habib et al., 2005, 2010), Go-ICP (Yang et al., 2013b), are representative approaches, which minimize distances between corresponding elements. Apart from ICP-based algorithms, normal distribution transform (NDT) (Biber and Straßer, 2003) is also a widely used method in the folder of fine registration. However, for the fine registration methods, proper initial transformation estimation are needed to avoid incorrect local optimum. Coarse registration are often conducted before fine registration to provide appropriate initial transformation estimation for fine registration. In this paper, we address the problem of coarse registration. In order to achieve coarse point cloud registration, two key steps are involved, including the estimation of correspondences and the calculation of transformation parameters, among which finding correspondences is requisite the whole process.

In the following, we mainly review coarse registration methods. A wide variety of literature has reported solutions for marker-less registration through the utilization of geometric characteristics. Generally, coarse registration approaches can be grouped into three fundamental classes conforming to the principles that they used: feature description-based, geometric constraint-based, and global information-based approaches.

2.1. Feature description-based registration

For feature description-based registration approaches, the corresponding pairs between point clouds are identified through retrieving features with the most substantial similarity. In this retrieving process, an appropriate feature description plays an important role, usually implemented by feature descriptors. Various feature descriptors have been demonstrated in many studies that are useful in the feature retrieving and matching. An eligible feature descriptor should have two core characteristics, namely, high descriptiveness and rotation-invariance. High descriptiveness ensures a discriminative description of geometric features for non-corresponding points and substantial similarity between features of corresponding points. Rotation-invariance guarantees the robustness of the generated features which should not be influenced by rigid transformation between point clouds. Renowned examples of feature descriptors include scale-invariant feature transform (SIFT) (Flitton et al., 2010), fast point feature histogram (FPFH) (Rusu et al., 2009), rotational projection statistics (RoPS) (Guo et al., 2013) and signature of histogram of orientations (SHOT) (Tombari et al., 2010). However, the performance of descriptors (i.e., SIFT), highly depends on the saliency of input points, which is selected by keypoint detectors like Harris 3D. The detection of key points will highly influence the performance of both candidate selection and feature extraction. Furthermore, the basic principle for achieving rotation in-variance mainly counts on the pose normalization. For instance, SIFT achieves rotation-invariance in feature extraction by orienting the local reference frame (LRF) axis to the gradients' dominant orientation. However, the orientation of LRF is easy to be influenced by noise and outliers. An alternative is to obtain the local geometry statistics, which is easy to implement and fast to compute. However, the critical problem is that this kind of features may encounter low descriptiveness. Additionally, features can also be extracted from geometric primitives that clustered from points, such as lines (Habib et al., 2005; Ge and Hu, 2020), curves (Yang and Zang, 2014), planes (Xiao et al., 2013), surfaces (Ge and Wunderlich, 2016). Thus, the accuracy of extracting these geometric candidates for registration, such as keypoints or primitives, is an importance factor that influences the registration results. Besides, artifacts may also be brought in when extracting geometric primitives, such as lines, planes, or surfaces.

2.2. Geometric constraint-based registration

Unlike feature description-based methods, some methods use a geometric constraint formed by points or primitives as an indicator for retrieving and matching correspondences. This type of methods follows a different registration scheme, in which specially designed combinations of points or primitives matter to the identification of corresponding points. This specially designed combination of points or primitives can create a constraint when searching for candidates pairs of points, which significantly increases the efficiency than a random matching test. 4-points congruent set (4PCS) (Aiger et al., 2008) and its variants such as Super4PCS (Mellado et al., 2014), keypoint-based 4PCS (K4PCS) (Theiler et al., 2014), and semantic keypoint-based 4PCS (SK4PCS) (Ge, 2017) are representative approaches following this strategy. In this type of methods, corresponding sets of congruent points are identified through utilizing the constraint of intersection ratios and selected as candidates for finding correspondences. In the affine transformation, intersection ratios of four points congruent sets consisting of two pairs of points are invariant. Thus, by filtering out all four points sets follow intersection ratios from a given four points sets in the target point cloud, we can reduce the number of candidates in the source point cloud. Compared with feature description-based registration, the geometric constraint-based methods have higher robustness to occlusions and unequal densities, since the geometric constraint can be built on a larger scale than the features extracted from a local context. Similarly, instead of points, using the combination of different kinds of primitives, for example, two pairs of planes (Chen et al., 2019), is also a compelling choice. The use of geometric primitives like planes can upgrade the robustness of the geometric features, as they reduce the DoF and are less sensitive to uneven points density and outliers (Xu et al., 2019). For example, the measured distances between points in the point-based 4PCS methods are more sensitive to noise compared to the primitive-based one. The volumetric 4PCS (V-4PCS) (Huang et al., 2017) is also a method under the framework of 4PCS, which extended the surface expression to volumetric ones and shows a promising improvement in computational efficiency.

2.3. Global information-based registration

In the aforementioned registration categories, local information are mainly utilized and generated from points themselves or clusters of primitives. Registration can also make use of global features derived from the entire point clouds. For instance, in the previously mentioned NDT-based methods, points were transformed into a normal distribution, the natural distribution of which forced alignment between point clouds (Magnusson et al., 2007). The distribution of point densities is another global indicator for alignment. In some representative methods, coherent point-drift (Myronenko and Song, 2010) and kernel affinity correlation (Tsin and Kanade, 2004) were applied on the density for finding correspondences. In a recent work of (Dong et al., 2018), global features were used for fast orientation of multi-scan unordered point clouds. In our previous work (Huang et al., 2019), 3D point clouds of a highly complicated scenario were projected into 1D histograms and 2D images for achieving registration in low-dimensional spaces. These projected histograms and images were also a global expression of original point clouds. Theoretically, the global information-based registration methods are more robust than the local feature-based ones, but large overlap ratios are usually required. Otherwise, the approaches based on global features may make a significant difference.

3. Principle of robust estimation of 3D transformation in the frequency domain

The core of a point cloud registration is the estimation of 3D transformation between two coordinate frames. In traditional point cloud registration methods, as we have mentioned, the registration relies heavily on matching correspondences via local geometric features. They firstly extract key points or feature points from both source and target data and then conduct the matching of corresponding points with features for estimating the transformation between different coordinate frames. Unlike conventional methods, in this paper, we proposed a new global information-based registration strategy following a principle which estimates 3D transformation in the frequency domain robustly. Following the proposed principle, our strategy converts the entire point clouds into 3D signals and regards them as global features. Then, the transformation between coordinate frames is achieved via the phase correlation in the frequency domain. Comparing with using local features of key points, the use of global features can increase the robustness. By transforming 3D points to 3D signals, we can separate and eliminate high-frequency parts representing noise and outliers, so that the matching of features could be more reliable. By using a novel robust and accurate phase correlation, the feature matching can be addressed by the optimization with a closed-form expression.

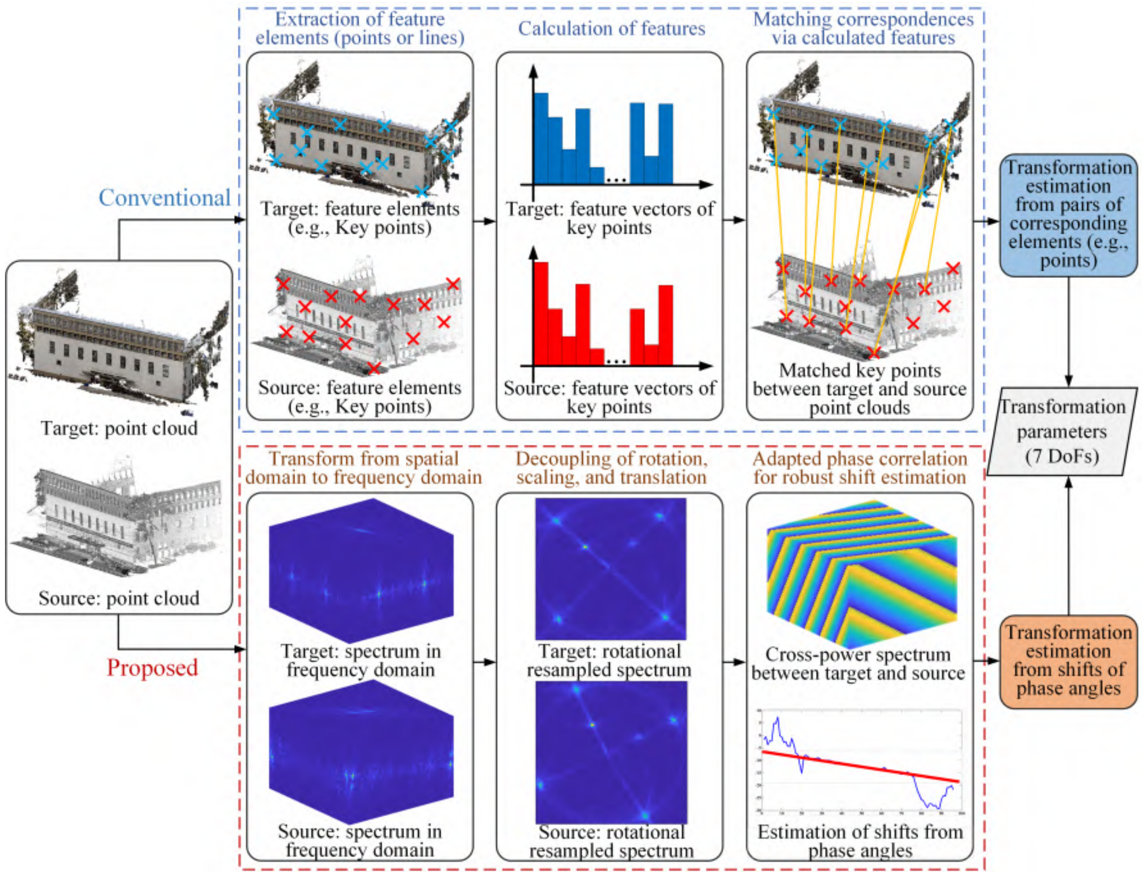

Figure 2: Comparison of workflows using our proposed principles and conventional feature description-based strategy.

In Fig. 2, we illustrate a comparison of workflows using our proposed strategy and conventional feature description-based ones. Specifically, the proposed principle for estimation of 3D transformation mainly comprises three principal aspects, including the transformation from the spatial to the frequency domain, decoupling of rotation, scaling, and translation, and robust and accurate shift estimation.

3.1. Transformation from the spatial domain to the frequency domain

The transformation from the spatial domain to the frequency domain is to create discrete 3D signals from unstructured and unordered points, which could be further used for the phase correlation. The transformation includes the voxelization and binarization of 3D points and 3D Fourier transformation of voxelized 3D data.

3.1.1. Voxelization and binarization of 3D points

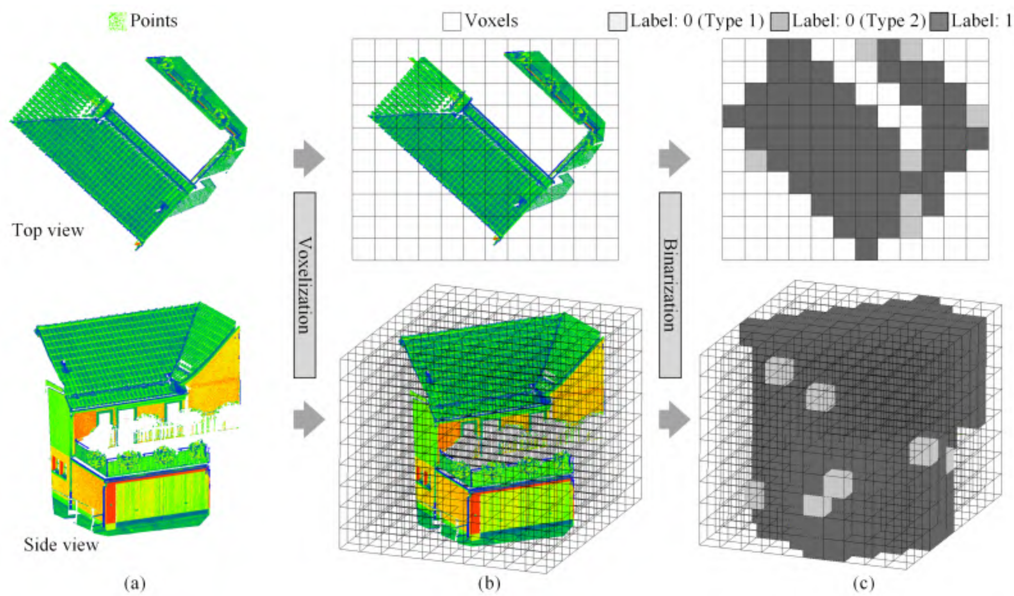

Figure 3: Voxelization and binarization of point clouds. (a) Original point clouds. (b) Voxelized point clouds. (c) Binarized voxels. It should be noted that Type 1 denotes the empty voxels annotated with value zero and Type 2 denotes the voxels with limited numbers of points and annotated with value zero.

Fig. 3 illustrates the voxelization and binarization of a given 3D point cloud. A voxelization process is presented to transform the unstructured and unordered points to a regularly resampled discrete 3D grid. Differing from the voxelization step in the other previous work, in which only the point cloud is voxelized, instead, the 3D space covering the entire point cloud is voxelized and resampled. The centers of all these voxels will be utilized to represent the point cloud and serve as basic input elements for the further process. Then, a binarization process is conducted on the resampled 3D grid, in which binary values (i.e., zero or one) are annotated to each voxel. The binary values actually denotes the occupancy of each voxel. It means that if points whose number is above a threshold fall into a voxel, the voxel will be marked as value one. Conversely, if a voxel contains limited number of points, it will be annotated with value zero. The threshold is actually set to filter out some isolated points. The thresholds are identified according to point densities of point clouds. In case that the bounding box of the point cloud is not a cubic, a zero-padding should be done, ensuring three dimensions are of the same sizes. The position and assigned labels of voxels will be used as inputs for further steps. In this way, an unevenly distributed point cloud can be resampled to a cubic grid, which represents the spatial distribution of point clouds and whose basic elements represent the corresponding point occupancy.

3.1.2. 3D Fourier transform of voxelized 3D data

In the previous step, the original point clouds have been transformed to regularly sampled discrete 3D signals. Assume that two signals are correlated to each other by shifts in the spatial domain denoted as $\mathbf{t}_s \in \mathcal{R}^{n\times 1}$, where n is the dimensions of the data. The correlation between the two signals can be expressed as:

$$s(\mathbf{x}) = r(\mathbf{x} - \mathbf{t}_s), \tag{1}$$

where $s(\mathbf{x})$ and $r(\mathbf{x})$ represent two signals in the spatial domain. Afterward, a fast discrete Fourier transform (FFT) can be conducted on these two signals to transform them from the spatial domain to the frequency domain:

$$\begin{cases} S(\mathbf{k}) = FFT(s(\mathbf{x})) \\ R(\mathbf{k}) = FFT(r(\mathbf{x})) \end{cases}, \tag{2}$$

where $S(\mathbf{k})$ and $R(\mathbf{k})$ are the corresponding Fourier transforms of $s(\mathbf{x})$ and $r(\mathbf{x})$. In this paper, we use lowercase letters to represent the spatial domain, while uppercase letters denote the frequency domain. If we carry out a phase correlation between $S(\mathbf{k})$ and $R(\mathbf{k})$, the relation between these two signals can be written as:

$$S(\mathbf{k}) = R(\mathbf{k})e^{-i2\pi(\mathbf{k}\mathbf{t}_s)}. \tag{3}$$

The normalized cross-power spectrum can be calculated as:

$$Q(\mathbf{k}) = \frac{S(\mathbf{k})R^*(\mathbf{k})}{|S(\mathbf{k})R^*(\mathbf{k})|} = e^{-i2\pi(\mathbf{k}\mathbf{t}_s)}, \tag{4}$$

where R^* represents the complex conjugate of R. The magnitude of Q is 1 after the normalization. From this equation, we can find that the translation $\mathbf{t}_s$ can be solved by exploiting the correlation between the signals. At this point, we have converted the estimation of translation in the spatial domain to an addressable problem in the frequency domain. This is a commonly used strategy in dense image matching in many previous works. However, when it comes to the point cloud registration, the problem is more complex, because the transformation between coordinate frames is an ill-posed problem of seven DoFs (Bellekens et al., 2015). For solving the ill-posed estimation problem of transformation, the transformation has to be decoupled and converted into a shift estimation task.

3.2. Decoupling of rotation, scaling and translation

The proposed strategy is to obtain the transformation by decomposing the transformation to several sub-problems, which can quickly solve shift estimation by phase correlation methods. First of all, we present the method used to decouple the transformation parameters, namely rotations, scaling, and translation. Before introducing details of the method, we present some basic concepts and notations used in the method. Assuming that two 3D voxel data can be presented as $f(\mathbf{x})$ and $h(\mathbf{x})$ which differ by rotations, translation, and scaling, the relation between the two 3D data can be expressed as:

$$h(\mathbf{x}) = sf(g(\alpha, \beta, \gamma)\mathbf{x} - \mathbf{t_s}), \tag{5}$$

in which $g(\alpha, \beta, \gamma)$ denotes rotations, s represents scaling factor, and $\mathbf{t_s} = [t_x, t_y, t_z]$ shifts the 3D voxel data by translation.

Two 3D voxelized discrete data can be transformed to the frequency domain using 3D FFT, then the relation between the spectrum of the data can be represented as follows:

$$H(\mathbf{k}) = s^3 F(g(\alpha, \beta, \gamma)\mathbf{k}s^{-1})e^{-i2\pi g(\alpha,\beta,\gamma)\mathbf{k}s\mathbf{t_s}}, \tag{6}$$

where $\mathbf{k} = [u, v, w]$ denotes the coordinates in frequency domain. It shows that the translation only has an impact on the phase of the spectrum. Thus, by calculating the magnitude of the spectrum, the 3D translation can be decoupled. The relation can be simplified as:

$$|H(\mathbf{k})| = s^3|F(g(\alpha, \beta, \gamma)\mathbf{k}s^{-1})|. \tag{7}$$

As shown by the equation, the spectral magnitude is influenced by a combination of rotations and scaling. A decoupling process is needed for estimating rotations and scaling separately. Explicitly, the rotations orient the 3D structure of the magnitude of the spectrum in the same way as it does for the original 3D data in the spatial domain. At the same time, the scaling affects the spectral magnitude in two aspects. One aspect is that the cubed scaling s^3 only influences the amplitude of the magnitude spectrum. However, the amplitude does not influence the structural information, indicating that it makes no difference in the phase matching procedure. Another is that the term s^{-1} indicates that the scale difference between signals in the spatial domain shows a reciprocal effect on the spectrum in the frequency domain. It means that scale also influences the structural information of the spectrum. Thus, in order to decouple rotations with scale, the spectral magnitude is radially accumulated. By accumulating spectral data radially, we can obtain a spherical function on which rotations present shifts of the structural information. Then, only rotations remain in terms of the accumulated spectrum. Thus, the general procedure of the seven DoFs transformation estimation is to estimate the rotations using the accumulated spectrum first and subsequently estimate other transformation parameters.

3.3. Adapted phase correlation for robust shift estimation

Via the use of phase correlation, we can convert the spatial translation estimation to the underlying shift estimation of phase angle differences. The main concept of phase correlation is that any shifts between two correlated signals (i.e., 2D images or 3D discrete signals) in the spatial domain can be represented as a phase shift in the frequency domain. Compared with correlation-based solutions which are also widely used, phase correlation seems to be more robust and accurate. Simultaneously, the processing efficiency is improved as a fringe benefit.

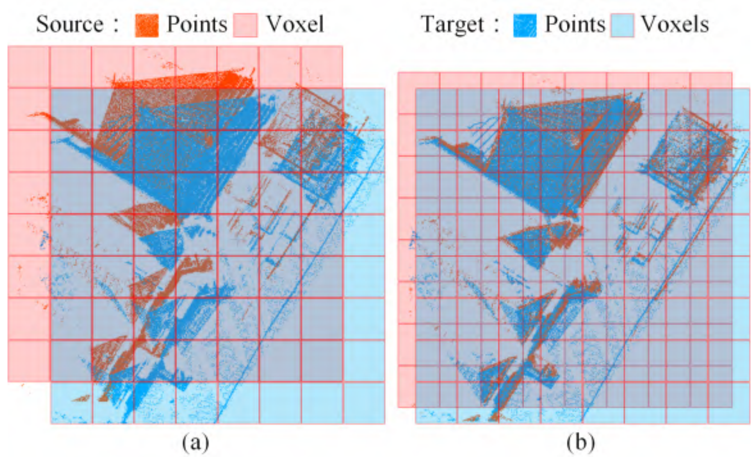

Figure 4: Registration (top view) with (a) voxel level accuracy and (b) sub-voxel level accuracy.

However, if we apply classic phase correlation methods (e.g., estimating shifts from the peak of the inversed FFT of the cross-power spectrum) to point cloud registration, we will encounter problems. For example, the estimated shifts can only achieve a voxel-level accuracy, which directly relates to the granularity of voxelization. In Fig. 4, we display a sketch showing a comparison between registrations with voxel- and subvoxel-level accuracies. To obtain an accurate registration, a sub-voxel level accuracy is mandatory, and this should be addressed by a fine-estimation of phase angle differences in the phase correlation. Moreover, as we have previously mentioned, the outliers and non-overlap areas will result in noise in the frequency domain signals, so we need to overcome these disturbances in the estimation of phase angle differences simultaneously. To this end, we proposed a novel multidimensional phase correlation method using merely the low-frequency components and $\ell 1$-normalized linear fitting for an accurate and robust shift estimation.

3.3.1. Multidimensional phase correlation

As assumed, the signals to be matched are in three dimensional, thus, the coordinates and shifts can be written as $\mathbf{k} = [u, v, w]$ and $\mathbf{t}_s = [t_x, t_y, t_z]$, respectively. It should be noted that although the solutions are provided in the 3D version, it can be easily adaptive to other multidimensional cases (i.e., 2D). In this case, the normalized cross-power spectrum can be written as:

$$Q(u, v, w) = e^{-i2\pi(ut_x + vt_y + wt_z)}. \tag{8}$$

The inverse Fourier transform (IFT) of $Q(\mathbf{k})$ contains a Dirac delta function in an

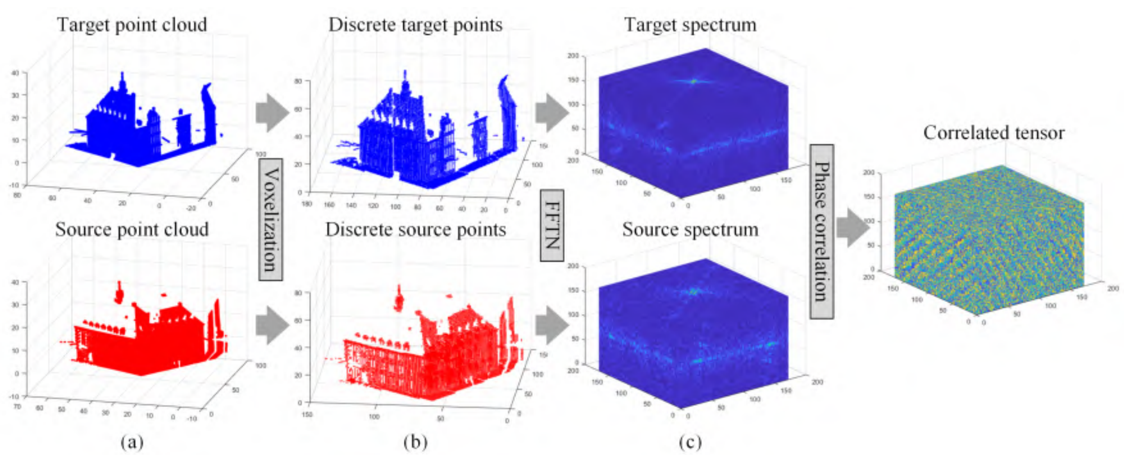

Figure 5: Multidimensional phase correlation. (a) Original point clouds. (b) Voxelized 3D points. (c) Spectrum of discrete 3D signals after FFT. (d) Correlated tensor from phase correlation of spectrums.

analytical way. Thus, the phase correlation result can be obtained by finding the Dirac peak, whose coordinates corresponds to the estimated parameters. However, this solution has a two-fold drawback. On the one hand, when the noise level is high, it will be hard to find a single peak for the function, which will lead to the failure or mistake in estimating the shifts. On the other hand, this kind of strategy is only able to produce a result in the accuracy of integer voxels or pixels, as shown in Fig. 4. This level of accuracy can not fulfill the requirement of registration of different scenes, especially for large-scale scenarios where the voxel size cannot be set as a small value.

Although there are some solutions proposed to improve the accuracy of the sub-pixel level by interpolation, the fitting of a high-dimensional polynomial function is not always robust, especially in high noisy cases. For tackling this problem, rather than sticking to interpolating the peak by some high-dimensional functions, many other methods have been reported aiming at improving the accuracy of phase matching to the sub-pixel level. An elegant way to solve the unknown shift parameters is to fit the phase difference angle, which can be represented as a linear function (i.e., 3D plane function). However, this is only feasible in an ideal situation. The real case is that noise, outliers, and the low-overlapping ratio of point clouds will produce strong disturbances to the cross power spectrum. Furthermore, the phase unwrapping of the high dimensional tensor will face an ill-posed problem with a high-level noisy cross-power spectrum tensor. In the following section, we will present our solution to the problems mentioned above.

3.3.2. Extraction of low-frequency components and signal decomposition

After obtaining fourier spectrum for each individual 3D signals and their corresponding normalized cross-power spectrum, it is of great importance to select from the frequency components and separate those low-frequency parts. Assuming that the 3D phase correlation between point clouds share similar characteristics to the 2D phase correlation between images, the same procedures can be conducted for the 3D signals. For 2D image matching, the concept is that high-frequency components corresponds to aliasing and noise, thus most of energy lies in the low-frequency components of the signals (Leprince et al., 2007). Thus, for the cross-power spectrum from the 3D phase correlation, a similar strategy is conducted to mask out around 80% of frequency components at the boundary of the tensor Q. Namely, only the center part of the tensor Q will be preserved for further processing. As for the esti-

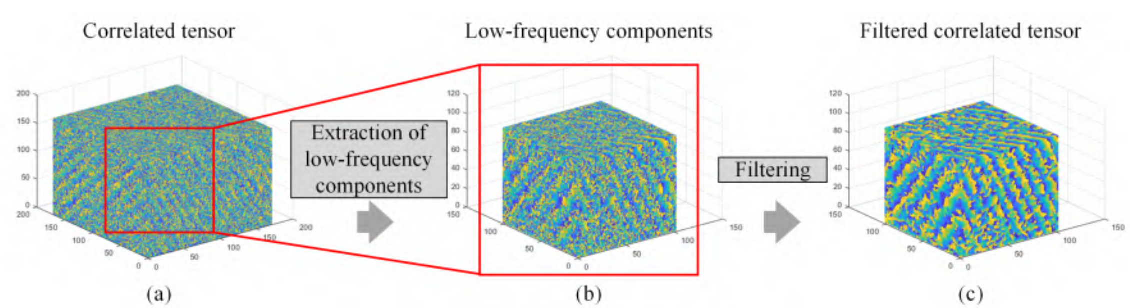

Figure 6: Illustration of the process of extracting low-frequency components. (a) Correlated tensor from phase correlation. (b) Extracted low-frequency components of the tensor. (c) Filtered low-frequency components of the tensor.

mation of the parameters, instead of fitting the high-dimensional plane, in this paper, a robust subpixel phase correlation method is applied, which combines the concept of SVD and $\ell 1$ normalization for robust estimation. The normalized cross-power spectrum can be rewitten as:

$$Q(u, v, w) = e^{-i(ut_x + vt_y + wt_z)} = e^{-iut_x} e^{-ivt_y} e^{-iwt_z} = Q_{x0}(u) Q_{y0}(v) Q_{z0}(w). \tag{9}$$

The cross-power spectrum can be represented by three rank-one signals. Thus, the task of the 3D shift estimation can be separated to several tasks which exploits the rank-one signals. Firstly, the SVD method can be utilized to divide the cross-power spectrum into several approximate rank-one signals. Thus, instead of solving phase wrapping and eliminating high-frequency components in high dimensions, these problems can be solved by finding 1D solution using the decomposed signals. Compared to the previous one, the 1D solution will be less sensitive to noise and outliers. Simultaneously, ill-posed problem of high-dimensional phase unwrapping can also be avoided. To calculate the coefficients of the fitted linear function of the decomposed and unwrapped signal, we adopt a robust algorithm in which $\ell 1$ normalization is utilized to add constraint and improve the model's robustness. Compared with $\ell 2$ normalization (e.g., least-squares adjustment), $\ell 1$ is less influenced by noise and outliers by adding constraints for the parameters.

3.3.3. Robust estimation of 3D shifts with $\ell 1$ norm

Although the low-frequency components in the cross-power spectrum are separated and extracted, Eq. 10 can still be utilized for the calculated of shift parameters. In order to estimate the parameters of this linear function, a robust estimator with $\ell 1$ normalization is adopted, which can be presented as follow:

$$\arg\min\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2 + \lambda\sum_{j=1}^{p}|\beta_i|\right), \tag{10}$$

where (x_i, y_i) are N pairs of data values of the decomposed signals and λ is a non-negative regularization parameter. In this problem, $\ell 1$ norm is involved, aiming to add constraints when estimating the linear function parameters. The alternating direction method of multipliers (ADMM) algorithm is utilized to solve the aforementioned optimization problem.

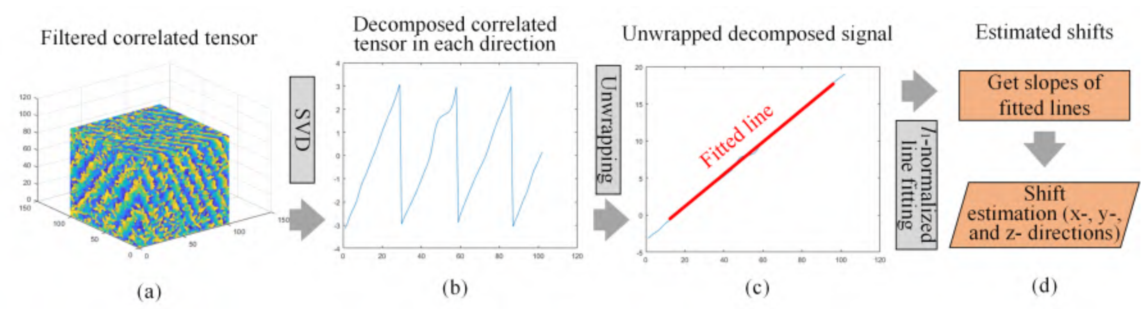

Figure 7: Illustration of the robust estimation of 3D shifts using $\ell 1$ norm. (a) Filtered low-frequency components. (b) Decomposed and wrapped 1D signal from low-frequency components. (c) Line fitting of unwrapped decomposed signal. (d) Estimation of shifts parameters.

Once the linear functions for each decomposed signals are estimated, parameters of the unwrapped phase angles of the identified components can be converted to the real estimated shift parameters:

$$\begin{cases} \Delta X = \delta x M/(2\pi) \\ \Delta Y = \delta y N/(2\pi) \\ \Delta Z = \delta z L/(2\pi) \end{cases}, \tag{11}$$

where M, N, L denote the dimensions of the input tensor, which are from the discrete Fourier transform (DFT). In the DFT that we used for transforming the point cloud into the frequency domain, the dimensions of the samples space in the frequency domain are $M \times N \times L$. Since the shifts are converted to the phase angle difference ranging from $-\pi$ to π, once we get the phase angle differences ΔX, ΔY, and ΔZ, we need to recover the real shifts by the use of Eq. 11 based on the sampled dimensions from DFT.

4. Application to the proposed GRPC method

Based on the proposed principle, we present our GRPC method for point cloud registration in the frequency domain, decoupling of transformation, and robust multidimensional phase correlation. Essential processing steps are summarized as a complete workflow shown in Fig. 8. In this workflow, the first step of the registration is the determination of rotations, which can be achieved by matching the accumulated spectrum in the Fourier domain, which is invariant to scaling and translation. Afterward, the scaling can be estimated using the rotationally aligned data by an adaptive Fourier-Mellin method. Finally, the 3D translation can be purely estimated by the shift estimation method, namely the robust 3D phase correlation by matching the 3D data, which has been re-scaled and re-rotated.

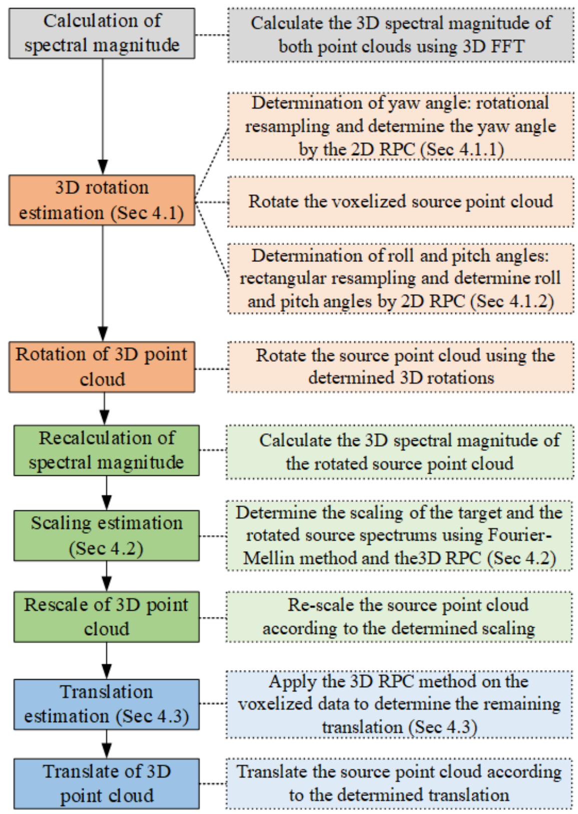

Figure 8: Detailed steps of the proposed GRPC method. Gray block stands for the transformation for the spatial domain to the frequency domain. Red blocks denotes the rotation estimation. Green blocks represent the scaling estimation. Blue blocks display the translation estimation. RPC denotes robust phase correlation.

Since the proposed GRPC method is under the framework of coarse registration, if more precise results are required, fine registration methods (i.e., ICP) can be conducted as a subsequent step to improve the registration accuracy.

4.1. 3D Rotation estimation

As presented in Section 3.2, rotations are presented as rotations of points on an accumulated spherical layer. In order to recover 3D rotations from the corresponding rotated spherical structure, in this paper, we aim at finding similar solutions to the way we use for the translation estimation, which is solved analytically. One solution is to use spherical harmonics. However, the main limitation is that the rotational information is recovered based on the standard cross-correlation, but the cross-correlation yields several peaks. There is a same problem as we have mentioned that even though the peak can be found, it is hard to achieve sub-voxel level interpolation since no closed-form way is provided. So the general idea is to resample the hemisphere of the accumulated spectrum. However, since the resampled layer is not an intrinsically 2D rectangular matrix, the structural distortions are dealt with a two-step strategy. First, the yaw angle is determined following the rotational behavior of the spherical structure. Then, the 3D spectrum is rotated according to the determined yaw angle. After the rotation, only roll and pitch angles remain their influence on the spherical structure. Thus, the remaining problem is to estimate roll and pitch by resampling the hemisphere in a rectangular way.

4.1.1. Determination of yaw angle

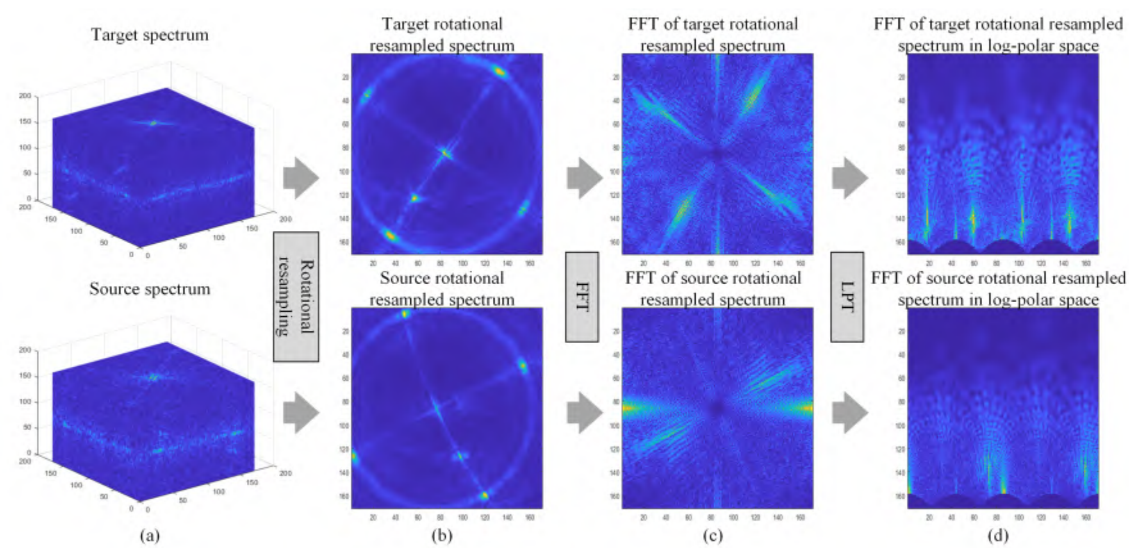

Figure 9: Illustration of the process of rotational resampling. (a) Spectrums from 3D signals. (b) Rotational resampled spectrum. (c) FFT of rotational resampled spectrum. (d) FFT of rotational resampled spectrum in log-polar space.

The general idea to determine the yaw angle is to treat it as a rotation of a resampled structure. The structure is resampled along with spherical coordinates. The accumulated spectrum can be expressed with a resampled spherical coordinate system. The coordinate system is as follows:

$$\begin{cases} v_i = 1, \dots N_{rot} \\ v_j = 1, \dots N_{rot} \\ \phi = arctan(\dfrac{v_i}{v_j}) \\ \theta = (v_i^2 + v_j^2)^{\frac{1}{2}} \dfrac{\pi}{N_{rot}} \end{cases}, \tag{12}$$

where N_{rot} denotes the size and v_i and v_j present the coordinates of the resampled images.

In accordance with the spectral magnitude, the spherical coordinates can be given as:

$$\begin{cases} u = rsin(\theta)cos(\phi) + \dfrac{N}{2} \\ v = rsin(\theta)sin(\phi) + \dfrac{N}{2} \\ w = rcos(\theta) + \dfrac{N}{2} \end{cases}, \tag{13}$$

$$f_{rot}(v_i, v_j) = \sum_{r=r_s}^{r=r_e} F(u, v, w) \tag{14}$$

In the resampled matrix, roll and pitch are shown as undesirable interference, which displays roughly as shifts between the matrices in x- and y-directions.

In order to recover the rotation between the resampled images, translation can be decoupled by calculating the Fourier magnitude spectrum. Then, the estimation of rotation can be transformed to a shift estimation problem by log-polar transformation (LPT) as illustrated in Fig. 9, where the spectrum of the two signals can be expressed as:

$$|F(r, \theta)| = |G(r, \theta + \theta_0)|, \tag{15}$$

$$|F(\log r, \theta)| = |G(\log r, \theta + \theta_0)|. \tag{16}$$

It is clear that the rotation is converted to a shift between the two signals. Thus, by finding the shift (x_0, y_0) in the log-polar space, the rotation can be estimated:

$$\theta_0 = y_0. \tag{17}$$

4.1.2. Determination of roll and pitch angles

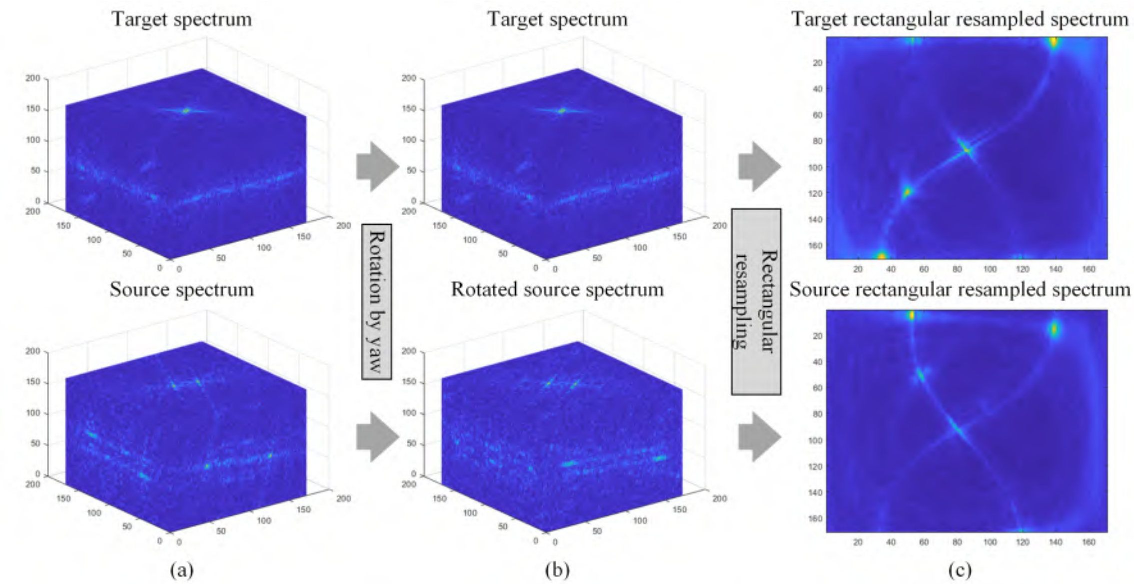

Figure 10: Illustration of the process of rectangular resampling. (a) Spectrums from 3D signals. (b) Target spectrum and rotated source spectrum. (c) Retangular resampled spectrum.

Different from the determination of yaw angle, roll and pitch angles are estimated simultaneously. First, the 3D voxel data is rotated based on the formerly estimated yaw. Then, the rotated spectral magnitude is attained using the same step, as mentioned before. For determining roll and pitch, the spectrum is re-sampled in a rectangular way by a perpendicular projection of the hemisphere into a matrix. Shifts between matrixes can roughly represent the roll and pitch. The resampled coordinate system can be expressed as:

$$\begin{cases} \gamma = -\frac{\pi}{2}(\frac{v_k - N_{rect}/2}{N_{rect}/2}), v_k = 1, \dots N_{rect} \\ \psi = \frac{\pi}{2}(\frac{v_l - N_{rect}/2}{N_{rect}/2}), v_l = 1, \dots N_{rect} \end{cases}, \tag{18}$$

where N_{rect} is the square size of the rectangular resampled images.

Correspondingly, the related accumulated spectrum can be calculated as:

$$\begin{cases} u = r\sin(\gamma)\cos(\psi) + \dfrac{N}{2} \\ v = r\sin(\gamma) + \dfrac{N}{2} \\ w = r\cos(\gamma)\cos(\psi) + \dfrac{N}{2} \end{cases}, \tag{19}$$

$$f_{rect}(v_k, v_l) = \sum_{r=r_s}^{r=r_e} F(u, v, w). \tag{20}$$

By determine the shifts using the same phase correlation method, roll and pitch can then be determined. Since all rotation parameters have been determined, the spectrum can be re-rotated using the determined angles. Only the scaling remains to influence the structure information of the spectrum magnitude.

4.2. Scaling estimation using Fourier-Mellin transform

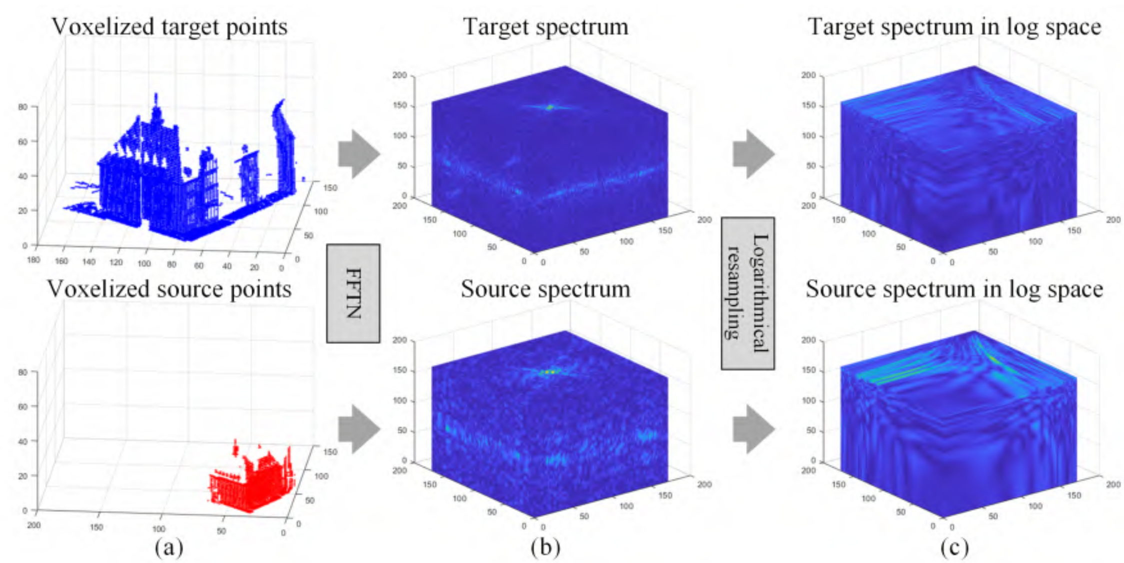

Figure 11: Scaling estimation using Fourier-Mellin transform. (a) Discrete 3D signals. (b) Spectrum of discrete 3D signals. (c) Spectrum in log space after FMT.

As mentioned in the previous sections, the radial accumulation of the spectral data is scale-invariant, which allows for the rotation-only registration. In this section, since the rotations have already been determined, Eq. 7 can be simplified as:

$$|R(\mathbf{k})| = \psi^3 |S(\mathbf{k}\psi^{-1})|. \tag{21}$$

The spectral magnitude can be transformed into a log space by Fourier Mellin transform, in which the Fourier magnitude spectrum is related to each other by:

$$|R(log(\mathbf{k}))| = \psi^3 |S(log(\mathbf{k}) - log(\psi))|, \tag{22}$$

which illustrates that spectral structure is logarithmically deformed along each direction. By taking the log transformation, the scaling is changed as a shift in each direction. Thus, by finding the shift (x_0, x_0, x_0) between the two spectra in the log space, the scaling factor can be estimated as:

$$\psi = e^{x_0}. \tag{23}$$

Note that there are common causes that the shifts in x-, y-, z-directions are different. However, under the assumption that we solve registration with seven DOF, the influence of scaling change on the spectrum's structure along each direction should be the same. Instead of adding a constraint, we apply an easy way under this situation. By estimating the 3D shift between the spectral structures in the log domain, the shifts in different directions can be determined. Thus, scaling for different directions can be easily calculated using Eq. 23. Subsequently, by finding the scaling that can produce the phase correlation's maximum peak, the scaling can be chosen among the three scaling factors.

4.3. Translation estimation using 3D phase correlation

Once the rotations and scaling have been determined, point clouds can be aligned according to the estimated parameters. Only translation remains. Thus, the further step is straightforward: the determination of 3D translation and can also be achieved simply by the 3D phase correlation. Without further procedures (i.e., transferring to other domains or conducting a resampling process), the translation can be directly determined by the proposed phase correlation method in the time domain using the aligned voxelized data. Assume that the estimated shifts are calculated as $(\Delta X, \Delta Y, \Delta Z)$. Afterward, considering the difference of the coordinates (X_0, Y_0, Z_0) calculated from the rough alignment, the estimated 3D translations should be (X_0+ΔX, Y_0+ΔY, Z_0+ΔZ), which are the final outputs.

5. Experiments

In order to test the efficiency and robustness of the proposed method, test datasets and evaluation criteria are required to evaluate the performance. Additionally, in order to test the versatility, three benchmark datasets with different point densities and different characteristics of scenes were selected for testing. In this section, we will introduce the test datasets and the evaluation criteria.

5.1. Test datasets

For evaluating the performance of the proposed registration method, experiments were conducted using three benchmark datasets, and their results were evaluated and analyzed. The first one is a pair of TLS point clouds from the ThermalMapper project acquired by the Jacobs University Bremen covering a large urban area (see Fig. 12). Table 1 shows the detailed information of the dataset, including the area size of the observed scene, the number of points, and the overlap ratios between scans. It should be noted that the target point cloud serves as a reference and the source point cloud is the one to be registered. The second dataset is a large-scale TLS point clouds registration benchmark (WHU-TLS) datasets published by Wuhan University (Dong et al., 2020), which provides multiview point clouds with varying point densities acquired from different scenes. We selected three representative scenes from the WHU-TLS dataset: a subway station, a park, and a cliff of a mountain. As shown by Fig. 13, point clouds acquired from these three different scenes show different geometric characteristics, which provide us valuable opportunities to test the strength and weakness of the proposed method. Detailed information for the selected point clouds is listed in Table 2. For this multiview dataset, the reference scan for each registration pair is also listed in Table 2. The last one is a set of scans from the Real-world Scans with Small Overlap (RESSO) dataset (see Fig. 14) (Chen et al., 2019). For this dataset, we used six TLS point clouds, which generated five registration test pairs. For each registration pair, Scan 2 is regarded as the reference scan. The detailed information of these scans is listed in Table 3. As shown by the table, the five pairs have different overlap ratios. By utilizing this dataset, the influence of different overlap ratios can be tested on the proposed method.

Table 1: Information of the Bremen TLS dataset.

Parameters	Target	Source
Area (m^2)	451 $\times$587	585 $\times$422
Number of points (million points)	15.2	15
Approxi. overlap ratio	0.85	

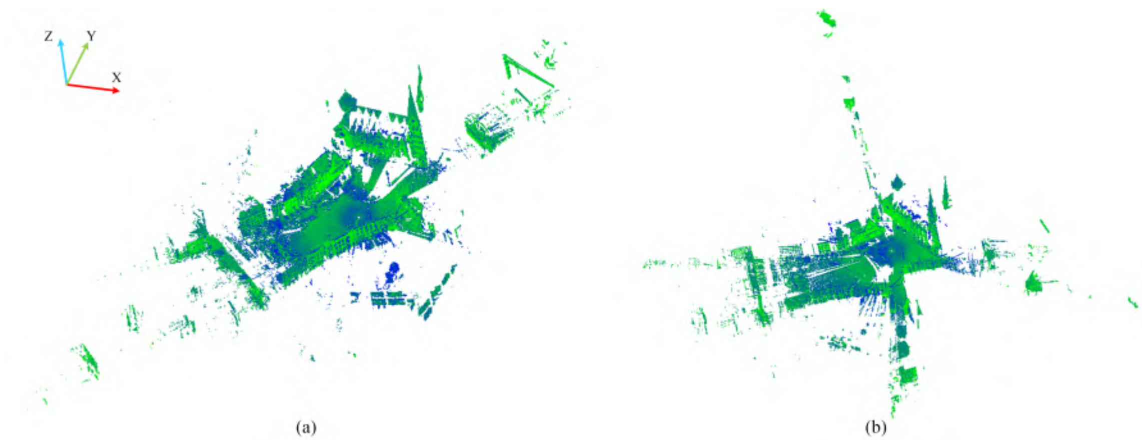

Figure 12: The Bremen TLS dataset. (a) Target and (b) source point clouds color-coded with intensities.

5.2. Evaluation criteria

The performance of the proposed method was evaluated from two different aspects. The first evaluation criterion is the registration accuracy. First, ground truth is needed for the evaluation of registration accuracy. For the first dataset, we manually aligned source and target point clouds, followed by an ICP refinement as ground truth. As a registration benchmark, the second and the third provided the accurately aligned source and target scans as ground truth. Then, the matching was performed between the source and target point clouds. The matching results can then be compared with the ground truth. The ground-truth transformation information of the two construction datasets was calculated based on the ground control information. The comparison between different algorithms was conducted using the rotation error e^r and the translation error e^t:

$$\Delta\mathbf{T} = \mathbf{T}_g(\mathbf{T}_r)^{-1} = \begin{bmatrix} \Delta\mathbf{R} & \Delta t \\ 0 & 1 \end{bmatrix}, \tag{24}$$

$$e^r = arccos(\frac{tr(\Delta\mathbf{R}) - 1}{2}), \tag{25}$$

$$e^t = \|\Delta t\|, \tag{26}$$

wherein $tr(\cdot)$ denotes the trace. Furthermore, $\mathbf{T}_g$ and $\mathbf{T}_r$ represent the transformation matrix of the ground truth and the estimated one, correspondingly. The second one is the time performance, which is used to test the efficiency of the proposed

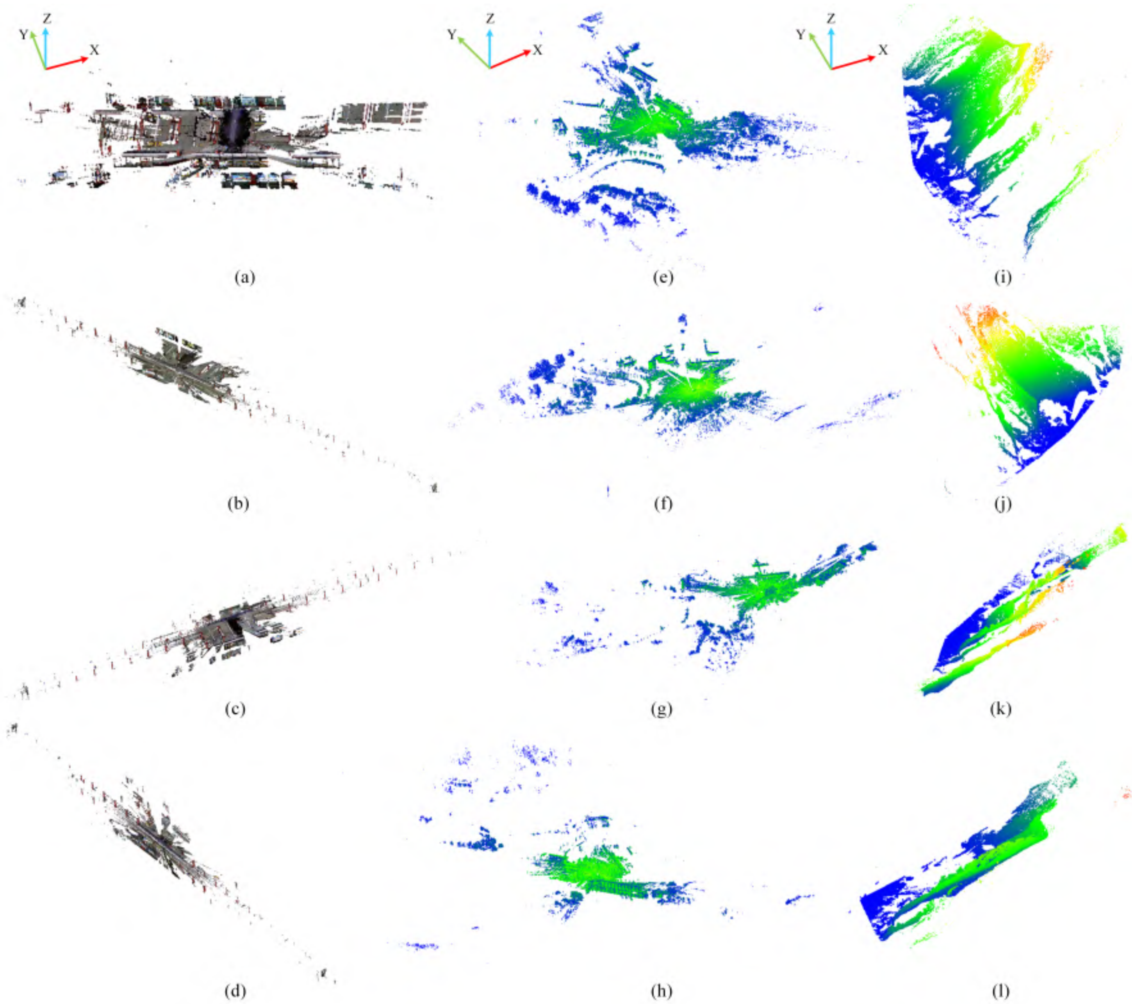

Figure 13: The WHU-TLS dataset. (a)-(d) are selected point clouds of the subway station dataset textured with RGB color. (e)-(f) are selected point clouds of the park dataset color-coded with intensities. (i)-(l) are selected point clouds of the mountain dataset color-coded with heights.

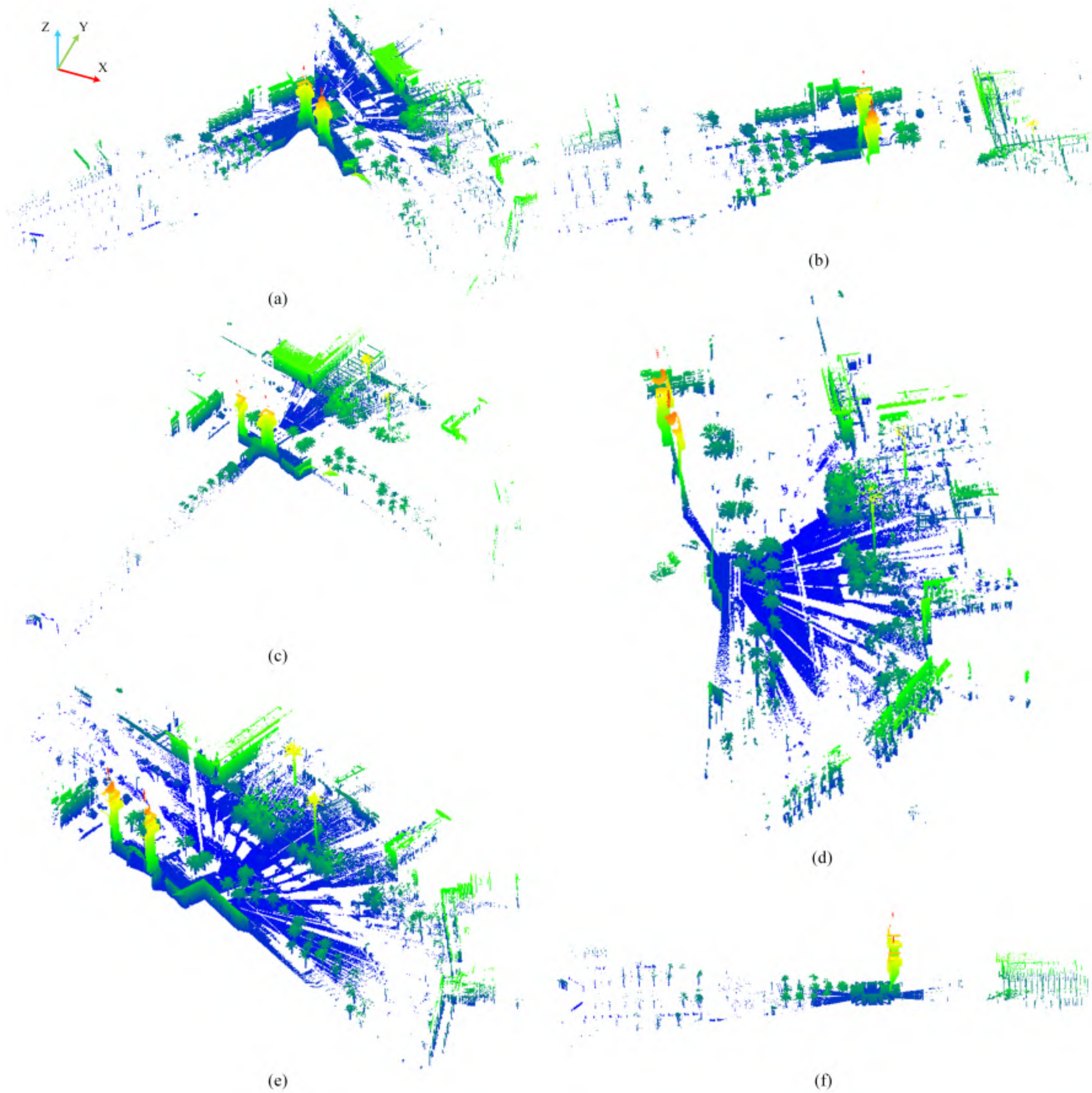

Figure 14: The selected scans from the Resso TLS dataset. (a) is the target point cloud, and (b)-(f) are source point clouds to be registered. All point clouds are color-coded with heights.

Table 2: Information of the WHU-TLS dataset.

Scene	Scan index	Area (m^2)	Number of points (million points)	Approxi. overlap ratio Reference scan index	ratio
Subway	5	73 ×60	41.7	/	/
	3	111 ×161	39.0	5	0.96
	4	203 ×35	39.1	5	0.92
	6	67 ×179	40.5	5	0.86
Park	14	461 ×564	3.9	/	/
	13	600 ×638	3.8	14	0.55
	15	526 ×552	4.9	14	0.47
	16	434 ×534	4.8	15	0.67
Mountain	4	255 ×277	3.4	/	/
	2	130 ×331	3.7	3	0.54
	3	122 ×346	3.5	4	0.50
	5	209 ×329	2.7	4	0.69

Table 3: Information of the Resso(7a) TLS dataset.

Parameters	Target	Source (a)	Source (b)	Source (c)	Source (d)	Source (e)
Scan number	2	1	3	5	6	7
Area (m^2)	275 ×280	195 ×273	256 ×258	231 ×192	218 ×183	177 ×260
Number of points (million points)	0.82	0.45	0.62	0.78	0.60	0.22
Approxi. overlap ratio	/	0.52	0.60	0.48	0.39	0.24

method. In our experiments, the execution time for the whole registration process was recorded. Our method was implemented using Matlab. All the experiments are conducted on a computer with an Intel i7-4710MQ CPU and 16GB RAM.

Additionally, our proposed GRPC method can also achieve an estimation of a seven DoFs transformation, including scaling and experiments, so evaluating the performance on scaling estimation is also conducted. The estimation of scaling is evaluated by scaling error:

$$\Delta s = |\frac{s_r}{s_g} - 1|, \tag{27}$$

where s_r and s_g are the scaling factor of ground truth and the estimated one, respectively.

6. Results

6.1. Results of Bremen dataset

The experimental results using the Bremen TLS dataset are listed in Table 4. In the experiments, the voxel size was set to 1 m. The filtering threshold for voxelization and binarization was set to 5.0. As shown in the table, the rotation error was about 0.04 degrees, and the translation error was nearly 0.25 m. In light of the requirement of coarse registration, our method's results are satisfactory. Additionally, the processing time was around 1 minute, which was efficient. Fig. 15 shows coarse registration results of the Bremen dataset. As illustrated in the figure, the source point cloud and the target point cloud were well aligned. It can be seen that the spires of the facade of the Bremen bank, and the walls were well matched.

In order to validate the effectiveness and efficiency of our proposed GRPC method, we selected several baseline methods for comparison, which were the method using Fast Point Feature Histograms(FPFH) (Rusu et al., 2009) and RANSAC process (FPFHSAC) (Holz et al., 2015), Keypoint-based 4-points congruent sets (K4PCS) (Theiler et al., 2014), and Voxel-based 4-plane congruent sets (V4PCS) (Xu et al., 2019). FPFASAC is a feature-based method, which combines FPFH features and a RANSAC process for estimating transformation parameters. K4PCS and V4PCS are both improved strategies in the framework of 4PCS. In K4PCS, keypoints are utilized to replace points in point clouds to reduce the number of candidates and improve the robustness of selected points. Differing from 4PCS and K4PCS, V4PCS replaces points by planes as candidates for the congruent pairs. The baseline results of these methods were provided in (Xu et al., 2019). As shown in the table, all registration methods provide acceptable results for a coarse registration. Compared with these baseline methods, our proposed GRPC method achieved the best registration accuracy considering both rotation and translation errors. Additionally, our proposed GRPC method also showed its superiority in its efficiency.

Table 4: Results of four registration methods using the Bremen dataset.

Methods	Rot_err (deg)	Trans_err (m)	Time (s)
FPFHSAC (Holz et al., 2015)	0.3601	0.0692	318
K4PCS (Theiler et al., 2014)	0.3682	0.4826	256
V4PCS (Xu et al., 2019)	0.1916	0.6312	78
Our GRPC	0.0453	0.2436	67

Figure 15: Registration result of the Bremen dataset using GRPC. (a) Source and target point clouds shown in the same coordinate frame. (b) Aligned source and target point clouds.

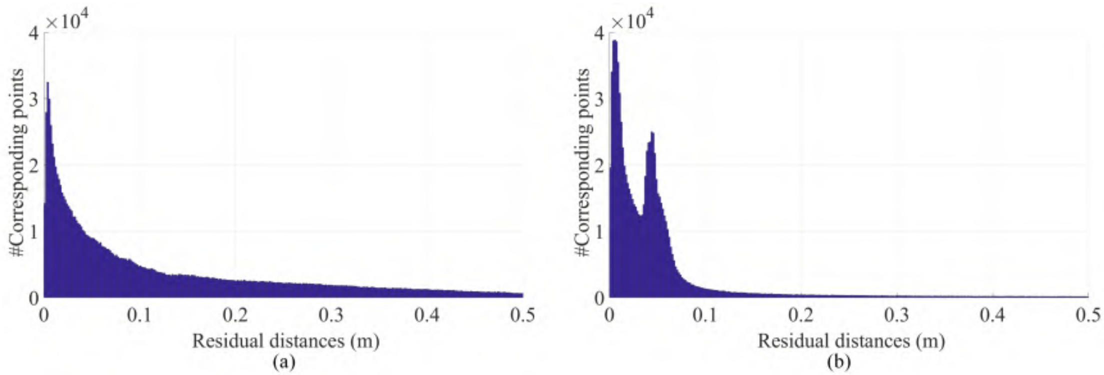

Figure 16: Histogram of residual distances between corresponding points in Bremen dataset between (a) ground truth and the align source scan and between (b) target and the align source scans.

It should also be noted that since the ground truth was generated by manual alignment followed by an ICP refinement, we also evaluated the quality of the ground truth by calculating the residual distances between corresponding points between the aligned source and target point clouds. In Fig. 16, the histograms of residual distances between corresponding points in the alignment results using the given ground truth and the aligned source scans using our method are shown. It can be seen that regarding the residual distances, our proposed GRPC method provides better alignment results compared with ground truth, with smaller mean residual distances and lower standard deviations being obtained.

6.2. Results of WHU-TLS datasets

To further evaluate the versatility of the proposed method to different scenes, three different scenes were selected from the WHU-TLS benchmark dataset for testing, including both regular-shaped areas (i.e., urban areas) and irregular-shaped areas (i.e., mountain cliffs). Since multiple scans were acquired for each scene, four scans were selected for testing, and each scan was matched to the corresponding reference scan. The voxel sizes set for the scenes of the subway, the park, and the mountain were 1.0 m, 1.5 m, and 1.0 m, respectively. Additionally, the filtering thresholds for voxelization and binarization were set to 3.0, 5.0, and 2.0. Table 5 lists the registration results of our proposed GRPC method and the baseline results using Hierarchical merging based multiview registration (HMMR) (Dong et al., 2018) provided by the publisher of the WHU-TLS dataset. The baseline method is also a hybrid method combining both global (for initial orientation) and local features (for fine registration) (Dong et al., 2020). As shown in Table 5, for the scene of the subway, the rotation errors of our proposed GRPC method were less than 0.2 degrees, and the translation errors were less than 0.6 m. Meanwhile, the processing time was less than two minutes. Compared with the baseline method, our GRPC provided better registration outputs in several cases (i.e., the matching between Scans 5 and 3 and Scans 5 and 6), with better results achieved in both rotations and translations.

Table 5: Performance comparison of our method and the baseline using the WHU-TLS dataset.

Scene	Registration pair (Target & Source)	Baseline (Dong et al., 2018) Rot_err (deg)	Baseline (Dong et al., 2018) Trans_err (m)	Our GRPC Rot_err (deg)	Our GRPC Trans_err (m)	Our GRPC Time (s)
Subway	5 & 3	Failed		0.0490	0.4848	65
	5 & 4	0.0722	0.7025	0.1841	0.2125	120
	5 & 6	0.0931	1.0286	0.0692	0.5493	93
Park	14 & 13	0.0864	0.0438	0.0795	0.4059	158
	14 & 15	0.0572	0.0358	0.0646	0.3202	137
	15 & 16	0.0256	0.0112	0.0862	0.7704	135
Mountain	3 & 2	0.0495	0.0180	0.2338	0.4010	79
	4 & 3	0.0422	0.0090	0.1827	0.2946	74
	4 & 5	11.1691	7.9453	0.1332	0.3263	69

For the scene of the park, our proposed GRPC method achieved less than 0.1 degrees rotation errors, which was the same level as the results provided by the baseline method. However, the proposed method's translation errors were larger than 0.3 m, while the baseline method can provide translation errors at a centimeter level. Compared with the first scene, the processing time is longer, about two and a half minutes. For the scene of the mountain, the proposed method provided results about a rotation error of around 0.2 degrees and a translation error of about 0.3

m. In general, compared with our proposed method, baseline (Dong et al., 2018) achieved better results in most cases with rotation errors less than 0.1 degrees and most of the translation errors in centimeter-level, which may benefit from the iterative optimization procedure. However, it also failed in some cases, namely, the matching of Scans 4 and 5 of the mountain scene when details of scans changed in a broad range. It shows that one advantage of our GRPC is that it runs stable under different situations in various registration datasets. Additionally, our proposed method is efficient concerning the processing time. Fig. 17 depicts the registration results of multiple scans in different scenes from the WHU-TLS dataset. The reference scans for the multiscan registration in the scene of the subway, the park, and the mountain were Scan 5, Scan 14, and Scan 4, respectively. As illustrated in the figure, it can be observed that walls, buildings in the park, and the mountain's valley were well aligned.

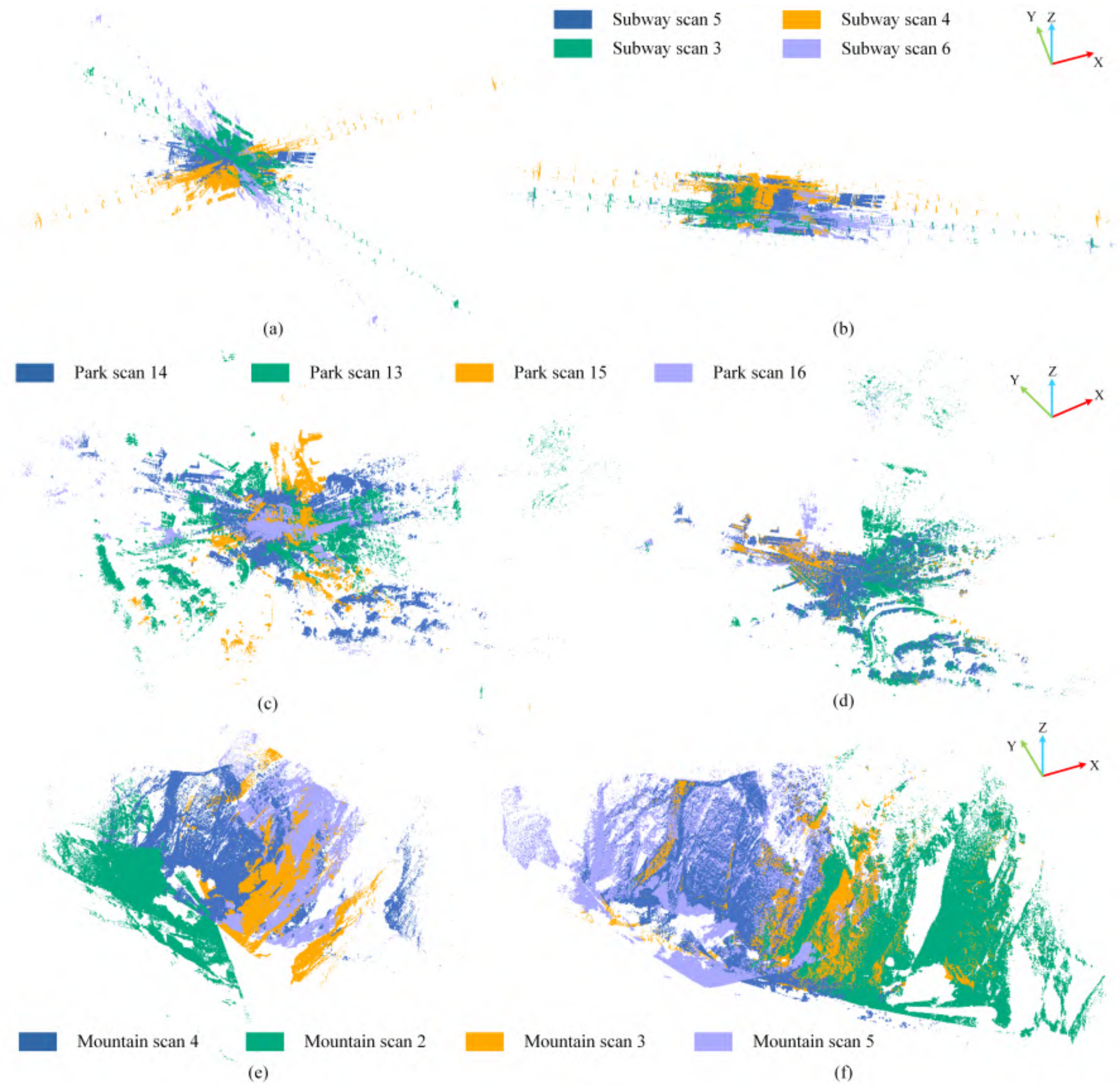

Figure 17: Registration result of the WHU-TLS dataset using GRPC, with color representing different scans. (a), (c), and (e) Source and target point clouds shown in the same coordinate frame. (b), (d), and (f) Aligned source and target point clouds.

6.3. Results of RESSO datasets

Apart from the Bremen-TLS and WHU-TLS datasets, we further tested our proposed GRPC method using another benchmark dataset, namely the Resso dataset. In the experiments, the voxel size was set to 1.0 m. The filtering threshold for voxelization and binarization was set to 3.0. In Table 6, it can be seen that the rotation errors were all smaller than 0.3 degree, and the translation errors were less than 0.6 m. Besides, the processing is comparatively fast, with processing time less than 50 s. The registration results of the baseline method, Plane-based descriptor (PLADE) (Chen et al., 2019), provided by the data publisher, are also given in Table 6. In PLADE, a plane and line-based descriptor are utilized to establish correspondences between point clouds. It can be seen that our proposed GRPC method always performed better in estimating rotations. However, as for the estimation of translations, our GRPC and PLADE achieved almost the same level results.

Table 6: Performance comparison of our method and the baseline using the Resso dataset.

Scene	Registration pair	Baseline (Chen et al., 2019)		Our GRPC		
	(Target & Source)	Rot_err (deg)	Trans_err (m)	Rot_err (deg)	Trans_err (m)	Time (s)
7a	2 & 1	0.3265	0.2082	0.2650	0.5610	45
	2 & 3	0.0810	0.0854	0.0727	0.3075	46
	2 & 5	0.4475	0.3626	0.1951	0.4000	43
	2 & 6	0.5060	0.5741	0.0485	0.2909	48
	2 & 7	0.2497	0.4057	0.2844	0.2391	44

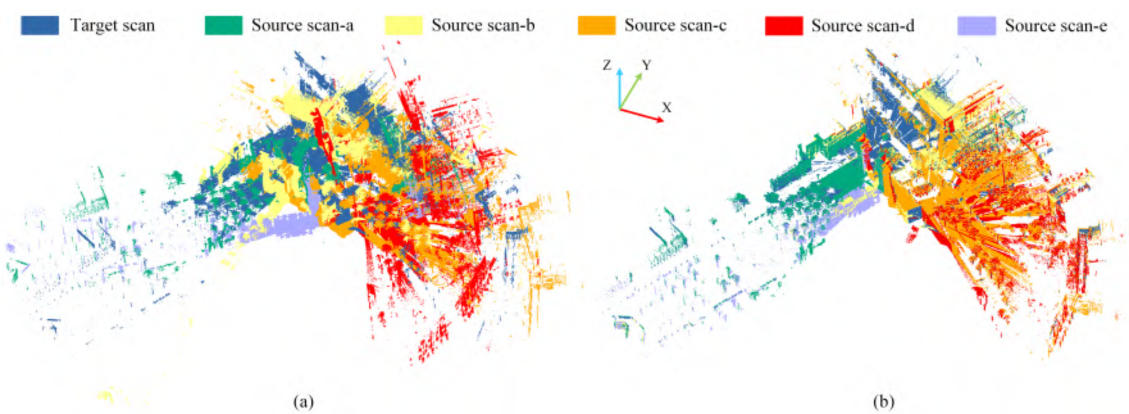

Figure 18: Registration result of the Resso dataset using GRPC, with color representing different scans. (a) Source and target point clouds shown in the same coordinate frame. (b) Aligned source and target point clouds.

The visualized results of the registration of the selected scene in the Resso dataset are shown in Fig. 18. It can be observed that the spires, palm trees, and walls of buildings were well matched.

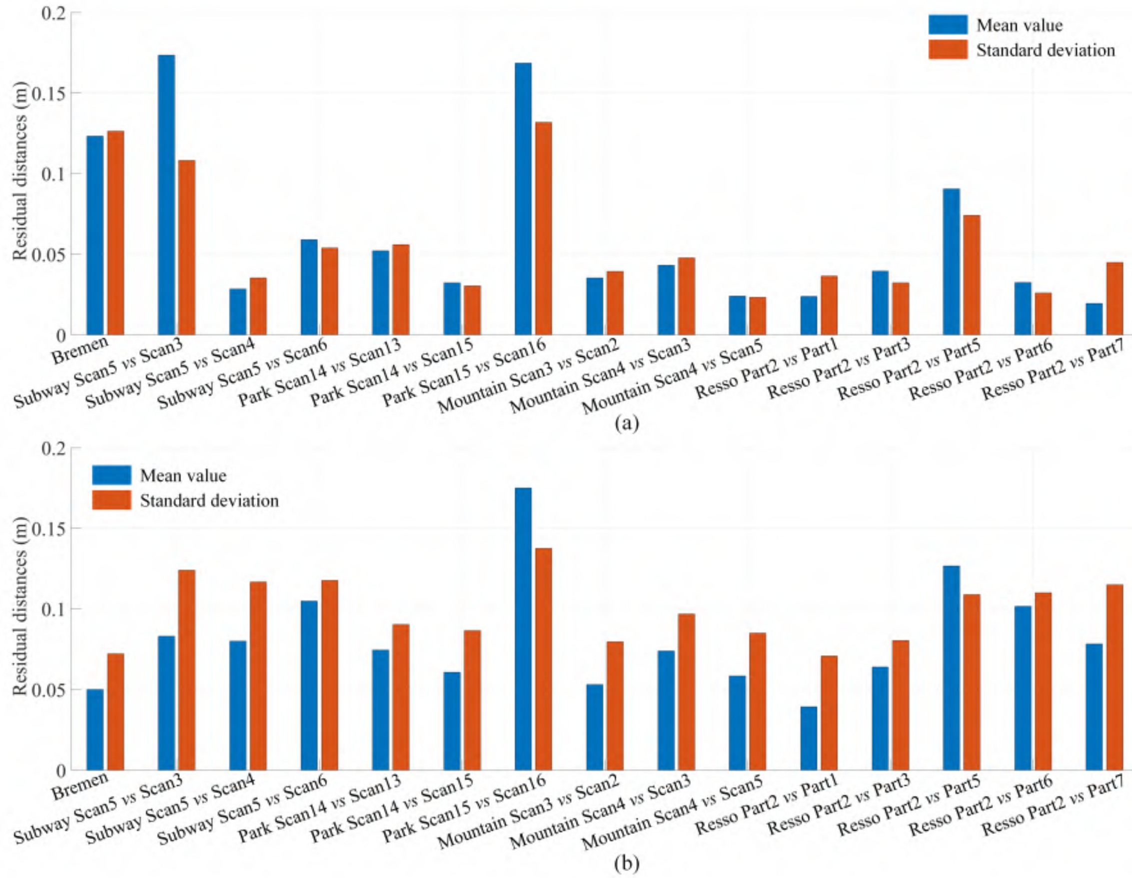

Figure 19: Mean values and standard deviation of residual distances between corresponding points in all pairs of scans between (a) ground truth and the align source scan and between (b) target and the align source scans.

7. Discussion

7.1. Influence of data properties

Three benchmark datasets for point cloud registration were tested in the experiments, including different point densities, different coverage areas, and different scenes. In Fig. 19, mean values and standard deviations of the residual distances between corresponding points in the aligned results are shown, in which both ground truth and aligned source scans using our proposed method were used as references.

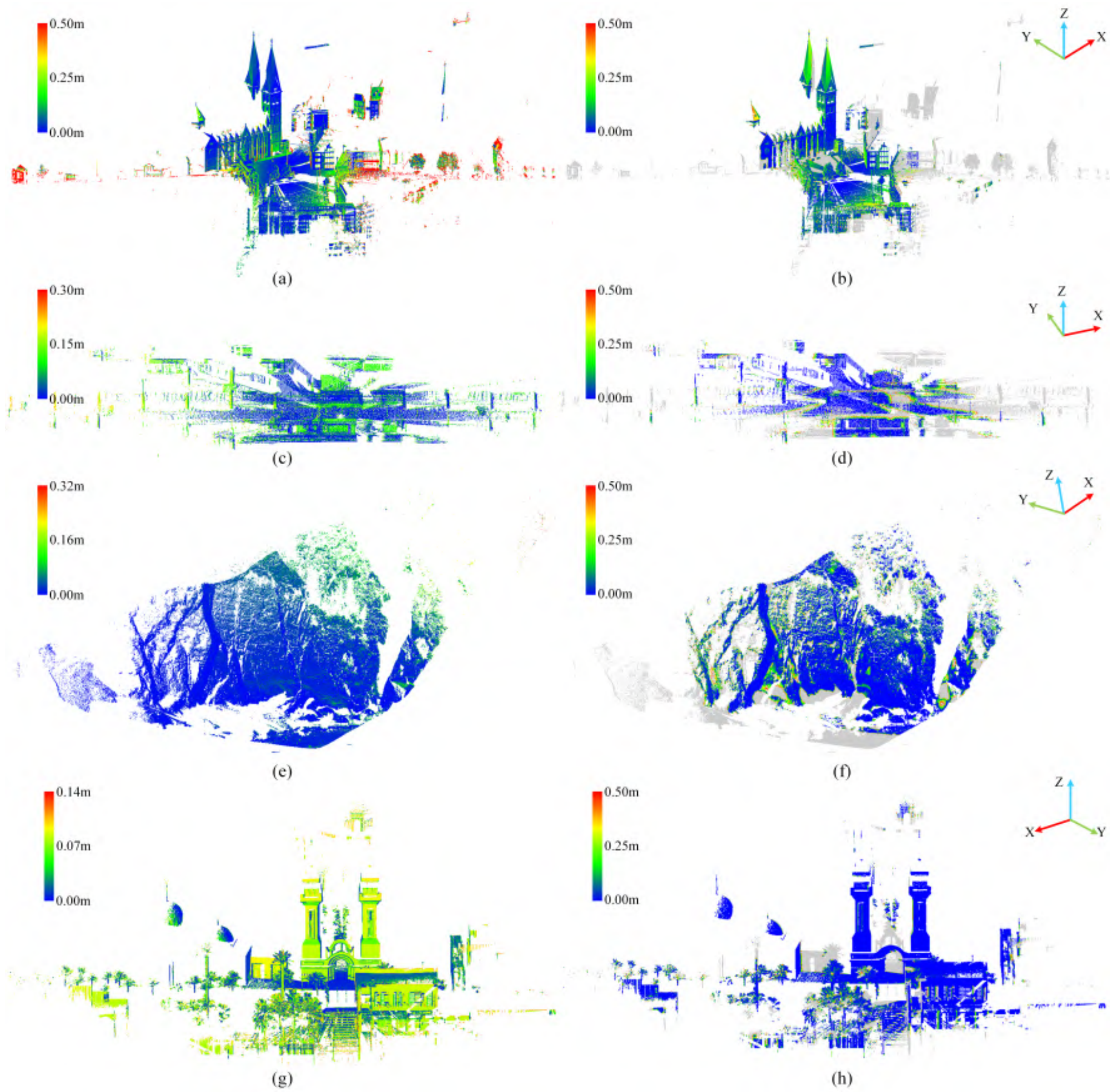

Figure 20: Selected registered results colored by the residual distances between corresponding points, where the gray points represent the non-overlap areas. Point distances between ground truth and the align source scan in (a) Bremen, (c) WHU-TLS subway, (e) WHU-TLS mountain, and (g) Resso. Point distances between target and the align source scans in (b) Bremen, (d) WHU-TLS subway, (f) WHU-TLS mountain, and (h) Resso.

For most registration pairs, the mean values and standard deviations of residual distances in results using our proposed method were close to those using ground truth. It means that our proposed GRPC method can provide acceptable results under different evaluation criteria, even employing checking point-by-point details. Additionally, we selected several representative registration pairs and illustrate the distribution of registration errors in Fig. 20. As shown in the figure, for most registration pairs, the distance errors were less than 0.25 m. By comparing different registration pairs, it can be seen that although the geometric characteristics of the acquired scenes changes and data property changes, our proposed GRPC method can always produce nearly equal and high quality of registration.

7.2. Influence of voxelization resolutions

The resolution of voxels is a significant factor influencing the result of registration. The resolution represents the geometric size of each voxel used in the step of voxelization and binarization. In the experiments, two registration pairs were selected from the aforementioned tested datasets. The first one is the pair of Scans 2 and 3 from the Resso dataset, which serves as a representative of regular-shaped areas. The other one is the pair of Scans 3 and 2 from the scene of a mountain cliff in the WHU-TLS dataset, which stands for a representative of irregular-shaped areas. In the experiments, the sizes of voxels range from 1.0 m to 3.0 m with a progressively increasing rate of 1.0 m per test. In Fig. 21, the registration results, including rotation errors, translation errors, and processing time, are provided.

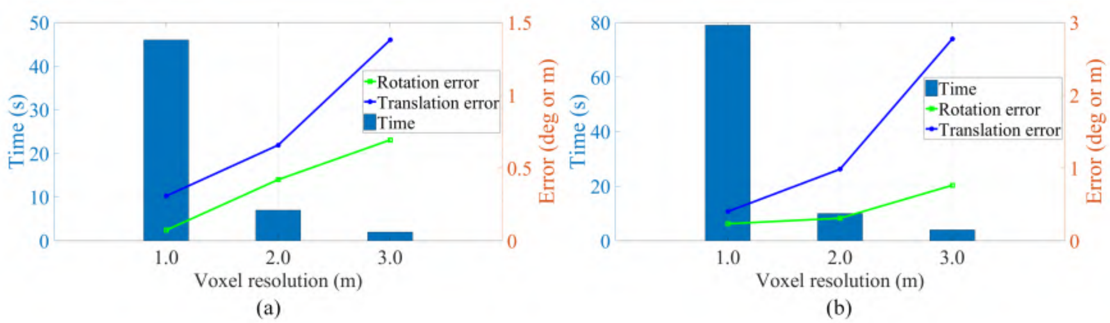

Figure 21: Sensitivity analysis on the resolution of voxel resolution. (a) Results using Resso dataset. (b) Results using WHU-TLS mountain dataset.

As we can predict, when the voxel size gets larger, the execution time will decrease. The results perfectly proved this assumption. For both datasets, the processing time experienced a remarkable drop along with the increment of voxel resolution. On the other hand, it is also noticeable that both rotation errors and translation errors for the two datasets showed drastic improvements. For the Resso dataset, the rotation errors increased from less than 0.1 degrees to 0.7 degrees, and the translation errors rose from 0.3 m to larger than 1 m. For the WHU-TLS mountain dataset, the rotation errors increased from 0.2 degrees to about 0.8 degrees, while the translation errors expanded from 0.4 m to almost 2.8 m. It can be seen that no matter for regular-shaped areas or irregular-shaped areas, the voxel resolution is an essential factor that influences registration results. One reason may be that the voxel size actually defines the sampling rate in the process of voxelization. When the voxel size is large, a sparse samling is conducted on point clouds, which leads to strong aliasing effect.

7.3. Influence of scaling changes

All tested datasets we used in the experiments provide no scaling changes. To investigate the effectiveness of the scaling estimation and the influence on the estimation of other transformation parameters, we generated several simulated registration pairs of point clouds with scaling changes by zooming out the source point cloud. We selected the pair of Scans 2 and 3 from the Resso dataset and the pair of Scans 3 and 2 from the WHU-TLS mountain dataset as registration pairs for testing. The source scans, namely Scan 3 from the Resso and Scan 2 from the mountain dataset, were zoomed out with various scaling factors. The target scans remained no changes. As illustrated in Fig. 22, it is clear that when the scale difference gets larger, the registration accuracies decrease with larger rotation, translation, and scaling errors no matter for regular-shaped areas and irregular-shaped areas. It could be explained that when the point cloud is zoomed out with a large scale factor, the aliasing effect will be caused by a relatively sparse sampling process on the point cloud.

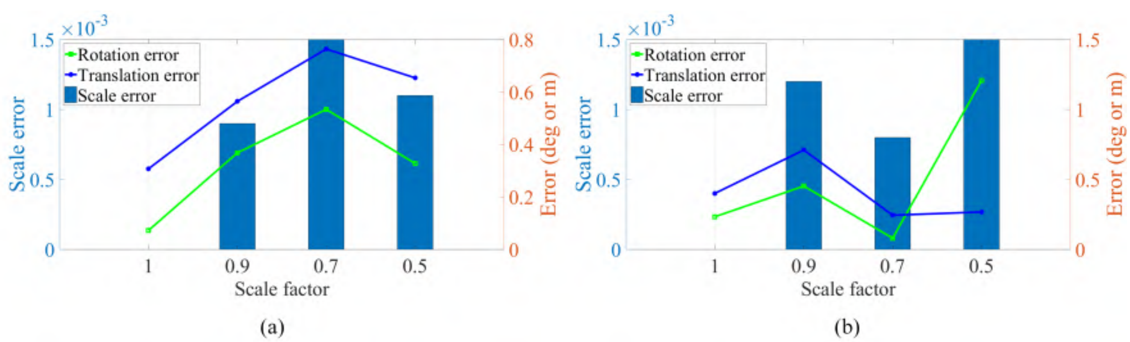

Figure 22: Sensitivity analysis on the changes of scale. (a) Results using Resso dataset. (b) Results using WHU-TLS mountain dataset.

Additionally, as shown in the figure, the influences of scaling changes on rotation errors and translation errors are almost with the same trend except for some individual cases. Since point clouds have been zoomed to approximately the same scale after scaling, the accuracy of estimated translations will not be influenced by aliasing effect caused by a sampling process but merely influenced by errors in the estimation of scaling and rotations.

7.4. Influence of signal-to-noise ratios

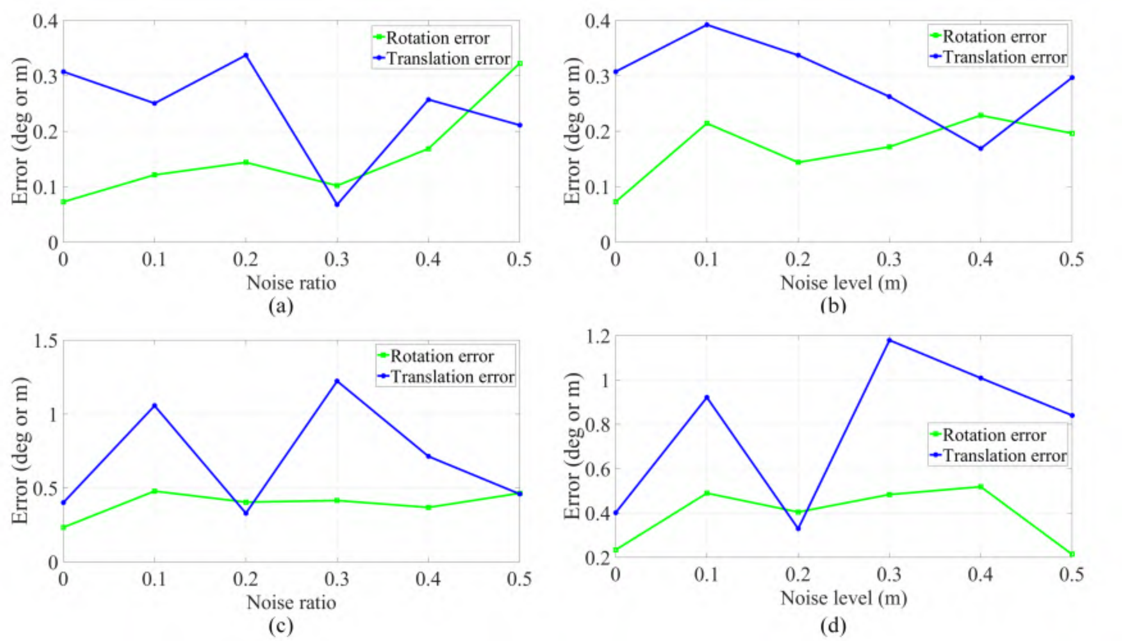

Figure 23: Rotation and translation errors with different noise ratios and noise levels using our proposed GRPC method. (a) and (b) Results using Resso dataset. (c) and (d) Results using WHU-TLS mountain dataset.

To validate the robustness of the proposed GRPC method, we conducted further experiments, which added noises to the original point clouds. In experiments, we selected two registration pairs from the aforementioned datasets. One pair is Scans 2 and 3 from the Resso dataset, and another pair is Scans 3 and 2 from the WHU-TLS mountain dataset. Meanwhile, Gaussian noises with different noise ratios and different noise levels were added to corresponding point clouds. It should be noted that noise ratio means the proportion of points that are changed to noise, while the noise level means the amplitude of the added noise. Thus, the influence of noise on regular-shaped and irregular-shaped datasets can also be investigated. The voxel size and the filtering threshold for voxelization and binarization were set as 1.0 m

and 3.0. As shown in Fig. 23, the registration accuracies vary at an acceptable level with changes in noise ratios and noise levels. It demonstrates the robustness of the proposed registration method and proves that the proposed method can still be effective in a highly-noisy situation. Comparatively, the registration of the mountain dataset is more sensitive to the influence of noise, with higher translation errors gained. However, for the mountain dataset, the estimation of rotations seems to be more stable under the changes of both noise ratios and noise levels.

7.5. Influence of different overlap ratios

In the real world, occasionally, it is unpredictable for a pair of scans to have varying overlap ratios, which is a challenging work for point cloud registration. Thus, we also investigated the influence of overlap ratios on the registration results using the proposed Go-PRC method. As depicted in Table 3, the dataset provides several scans with different overlap ratios varying from 0.60 to 0.24, but with the same data quality. The voxel sizes were all set to 1.0 m, and the filtering thresholds were set to 3.0, as mentioned in Section 6.3. In Fig. 24, overlap ratios, and their corresponding rotation and translation errors are shown. As illustrated in Fig. 24, the registration results are not directly positively influenced by overlap ratios. When the overlap ratio dropped from 0.52 to 0.24, the registration accuracy was still acceptable with a rotation error by about 0.3 degree and a translation error by 0.25 m. Generally, it shows that the proposed method is kind of robust to the variations of overlap ratios.

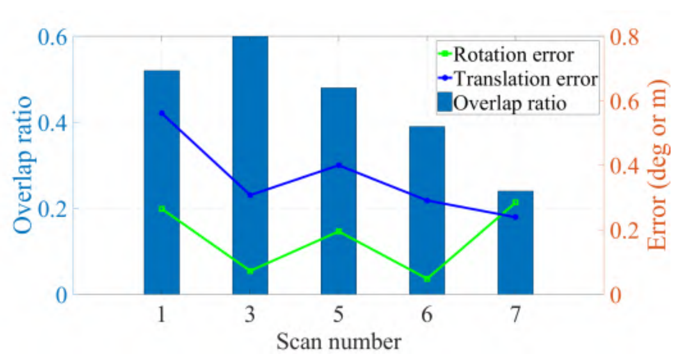

Figure 24: The rotation and translation errors of the registration results using Resso with different overlap ratios

8. Conclusion

In this paper, we propose a marker-free method called GRPC, which utilizes global features for efficient and robust registration of point clouds. The proposed

GRPC method converts the estimation of rotations, scaling, and translations to a problem of matching low-frequency components in the frequency domain. Specifically, the estimation of rotations, scaling, and translations is converted to a sequence of shift estimation tasks by a sequence of operations, including Fourier transform, resampling strategies, and Fourier-Mellin transform. Accurate estimations of shifts can be sequentially achieved by fitting the decomposed cross-power spectrum of global signal tensors in the Fourier domain using a robust estimator with $\ell 1$-norm. Experiment results using three TLS datasets from different sources and representing different scenes reveal that the proposed method is practical and efficient under different scenarios. Promising results also prove the versatility of the proposed method to different datasets with regular-shaped or irregular-shaped geometric characteristics. The proposed method can efficiently achieve registration with majority rotation and translation errors, reaching less than 0.2 degrees and 0.5 m and outperform state-of-the-art methods on the Bremen dataset and the baseline method on the majority scan pairs of the Resso dataset (in the tested scene). As for the WHU-TLS dataset, although in terms of registration accuracy, the baseline method outperforms our proposed method, our method can produce more stable results of satisfying quality even under significant changes in the scene. Additionally, it is also proved by the experiments that our proposed GRPC method is kind of robust to noise and is still effective and efficient under low-overlapping cases. Although several datasets covering various scenes were tested in our experiments, they were all acquired via TLS, with similar data characteristics. In our future work, the potential of utilizing global features in the frequency domain in the registration of a multisource dataset can be investigated.

Acknowledgement

This work was carried out within the frame of Leonhard Obermeyer Center (LOC) at Technical University of Munich (TUM) [www.loc.tum.de]. The authors would like to appreciate the technical support by Dr. Zhen Ye from Photogrammetry and Remote Sensing of TUM.

References

Aiger, D., Mitra, N. J., Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface registration. ACM Transactions on Graphics (TOG) 27 (3), 85.

Bae, K.-H., Lichti, D. D., 2008. A method for automated registration of unorganised point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 63 (1), 36–54.

Bellekens, B., Spruyt, V., Berkvens, R., Penne, R., Weyn, M., 2015. A benchmark survey of rigid 3d point cloud registration algorithms. International Journal on Advances in Intelligent Systems 8, 118–127.

Besl, P. J., McKay, N. D., 1992. Method for registration of 3-d shapes. In: Robotics-DL tentative. International Society for Optics and Photonics, pp. 586–606.

Biber, P., Straßer, W., 2003. The normal distributions transform: A new approach to laser scan matching. In: Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453). Vol. 3. IEEE, pp. 2743–2748.

Bosché, F., Ahmed, M., Turkan, Y., Haas, C., Haas, R., 2015. The value of integrating scan-to-bim and scan-vs-bim techniques for construction monitoring using laser scanning and bim: The case of cylindrical mep components. Automation in Construction 49, 201–213.

Chen, S., Nan, L., Xia, R., Zhao, J., Wonka, P., 2019. Plade: A plane-based descriptor for point cloud registration with small overlap. IEEE Transactions on Geoscience and Remote Sensing.

Dong, Z., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppäb, J., et al., 2020. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. ISPRS Journal of Photogrammetry and Remote Sensing 163, 327–342.

Dong, Z., Yang, B., Liang, F., Huang, R., Scherer, S., 2018. Hierarchical registration of unordered tls point clouds based on binary shape context descriptor. ISPRS Journal of Photogrammetry and Remote Sensing 144, 61 – 79.

Flitton, G. T., Breckon, T. P., Bouallagu, N. M., 2010. Object recognition using 3d sift in complex ct volumes. In: BMVC. No. 1. pp. 1–12.

Ge, X., 2017. Automatic markerless registration of point clouds with semantic-keypoint-based 4-points congruent sets. ISPRS Journal of Photogrammetry and Remote Sensing 130, 344–357.

Ge, X., Hu, H., 2020. Object-based incremental registration of terrestrial point clouds in an urban environment. ISPRS Journal of Photogrammetry and Remote Sensing 161, 218–232.

Ge, X., Wunderlich, T., 2016. Surface-based matching of 3d point clouds with variable coordinates in source and target system. ISPRS Journal of Photogrammetry and Remote Sensing 111, 1–12.

Gehrung, J., Hebel, M., Arens, M., Stilla, U., 2017. An approach to extract moving objects from mls data using a volumetric background representation. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4, 107.

Gressin, A., Mallet, C., Demantké, J., David, N., 2013. Towards 3d lidar point cloud registration improvement using optimal neighborhood knowledge. ISPRS Journal of Photogrammetry and Remote Sensing 79, 240–251.

Guo, Y., Sohel, F., Bennamoun, M., Lu, M., Wan, J., 2013. Rotational projection statistics for 3d local surface description and object recognition. International Journal of Computer Vision 105 (1), 63–86.

Habib, A., Detchev, I., Bang, K., 2010. A comparative analysis of two approaches for multiple-surface registration of irregular point clouds. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (1).

Habib, A., Ghanma, M., Morgan, M., Al-Ruzouq, R., 2005. Photogrammetric and lidar data registration using linear features. Photogrammetric Engineering & Remote Sensing 71 (6), 699–707.

Hebel, M., Arens, M., Stilla, U., 2013. Change detection in urban areas by object-based analysis and on-the-fly comparison of multi-view als data. ISPRS Journal of Photogrammetry and Remote Sensing 86, 52–64.

Hebel, M., Stilla, U., 2011. Simultaneous calibration of als systems and alignment of multiview lidar scans of urban areas. IEEE Transactions on Geoscience and Remote Sensing 50 (6), 2364–2379.

Holz, D., Ichim, A. E., Tombari, F., Rusu, R. B., Behnke, S., 2015. Registration with the point cloud library: A modular framework for aligning in 3-d. IEEE Robotics & Automation Magazine 22 (4), 110–124.

Huang, J., Kwok, T.-H., Zhou, C., 2017. V4pcs: Volumetric 4pcs algorithm for global registration. Journal of Mechanical Design 139 (11).

Huang, R., Xu, Y., Hoegner, L., Stilla, U., 2020a. Efficient estimation of 3d shifts between point clouds using low-frequency components of phase correlation. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences V-2-2020, 227–234.

Huang, R., Xu, Y., Hoegner, L., Stilla, U., 2020b. Temporal comparison of construction sites using photogrammetric point cloud sequences and robust phase correlation. Automation in Construction 117, 103247.

Huang, R., Xu, Y., Hong, D., Yao, W., Ghamisi, P., Stilla, U., 2020c. Deep point embedding for urban classification using als point clouds: A new perspective from local to global. ISPRS Journal of Photogrammetry and Remote Sensing 163, 62 – 81.

Huang, R., Ye, Z., Boerner, R., Yao, W., Xu, Y., Stilla, U., 2019. Fast pairwise coarse registration between point clouds of construction sites using 2d projection based phase correlation. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W13, 1015–1020.

Lafarge, F., Mallet, C., 2012. Creating large-scale city models from 3d-point clouds: a robust approach with hybrid representation. International Journal of Computer Vision 99 (1), 69–85.

Leprince, S., Barbot, S., Ayoub, F., Avouac, J.-P., 2007. Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. IEEE Transactions on Geoscience and Remote Sensing 45 (6), 1529–1558.

Magnusson, M., Lilienthal, A., Duckett, T., 2007. Scan registration for autonomous mining vehicles using 3d-ndt. Journal of Field Robotics 24 (10), 803–827.

Mellado, N., Aiger, D., Mitra, N. J., 2014. Super 4pcs fast global pointcloud registration via smart indexing. In: Computer Graphics Forum. Vol. 33. Wiley Online Library, pp. 205–215.

Myronenko, A., Song, X., 2010. Point set registration: Coherent point drift. IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (12), 2262–2275.

Polewski, P., Yao, W., Cao, L., Gao, S., 2019. Marker-free coregistration of uav and backpack lidar point clouds in forest ed areas. ISPRS Journal of Photogrammetry and Remote Sensing 147, 307 – 318.

Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. In: Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, pp. 3212–3217.

Theiler, P. W., Wegner, J. D., Schindler, K., 2014. Keypoint-based 4-points congruent sets–automated marker-less registration of laser scans. ISPRS Journal of Photogrammetry and Remote Sensing 96, 149–163.

Tombari, F., Salti, S., Di Stefano, L., 2010. Unique signatures of histograms for local surface description. In: European Conference on Computer Vision. Springer, pp. 356–369.

Tsin, Y., Kanade, T., 2004. A correlation-based approach to robust point set registration. In: European Conference on Computer Vision. Springer, pp. 558–569.

Tuttas, S., Braun, A., Borrmann, A., Stilla, U., 2017. Acquisition and consecutive registration of photogrammetric point clouds for construction progress monitoring using a 4d bim. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science 85 (1), 3–15.

Vosselman, G., Maas, H.-G., 2010. Airborne and terrestrial laser scanning. CRC Press.

Xiao, J., Adler, B., Zhang, J., Zhang, H., 2013. Planar segment based three-dimensional point cloud registration in outdoor environments. Journal of Field Robotics 30 (4), 552–582.

Xu, Y., Boerner, R., Yao, W., Hoegner, L., Stilla, U., 2017. Automated coarse registration of point clouds in 3d urban scenes using voxel based plane constraint. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W4, 185–191.

Xu, Y., Boerner, R., Yao, W., Hoegner, L., Stilla, U., 2019. Pairwise coarse registration of point clouds in urban scenes using voxel-based 4-planes congruent sets. ISPRS Journal of Photogrammetry and Remote Sensing 151, 106–123.

Yang, B., Dong, Z., Liang, F., Liu, Y., 2016. Automatic registration of large-scale urban scene point clouds based on semantic feature points. ISPRS Journal of Photogrammetry and Remote Sensing 113, 43–58.

Yang, B., Xu, W., Dong, Z., 2013a. Automated extraction of building outlines from airborne laser scanning point clouds. IEEE Geoscience and Remote Sensing Letters 10 (6), 1399–1403.

Yang, B., Zang, Y., 2014. Automated registration of dense terrestrial laser-scanning point clouds using curves. ISPRS Journal of Photogrammetry and Remote Sensing 95, 109–121.

Yang, J., Li, H., Jia, Y., 2013b. Go-icp: Solving 3d registration efficiently and globally optimally. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1457–1464.

Zhu, J., Gehrung, J., Huang, R., Borgmann, B., Sun, Z., Hoegner, L., Hebel, M., Xu, Y., Stilla, U., 2020. Tum-mls-2016: An annotated mobile lidar dataset of the tum city campus for semantic point cloud interpretation in urban areas. Remote Sensing 12 (11), 1875.