

Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark



Zhen Dong^a, Fuxun Liang^a, Bisheng Yang^{a,*}, Yusheng Xu^b, Yufu Zang^c, Jianping Li^a, Yuan Wang^a, Wenxia Dai^a, Hongchao Fan^{d,*}, Juha Hyppä^e, Uwe Stilla^b

^a State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

^b Photogrammetry and Remote Sensing, Technical University of Munich, Arcistr. 21, Munich 80333, Germany

^c School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science & Technology, 219 Ningliu Road, Nanjing 210044, China

^d Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway

^e Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, FGI, 02431 Masala, Finland

ARTICLE INFO

Keywords:

Benchmark data set
Deep learning
Point cloud
Registration
Terrestrial laser scanning

ABSTRACT

This study had two main aims: (1) to provide a comprehensive review of terrestrial laser scanner (TLS) point cloud registration methods and a better understanding of their strengths and weaknesses; and (2) to provide a large-scale benchmark data set (Wuhan University TLS: Whu-TLS) to support the development of cutting-edge TLS point cloud registration methods, especially deep learning-based methods. In particular, we first conducted a thorough review of TLS point cloud registration methods in terms of pairwise coarse registration, pairwise fine registration, and multiview registration, as well as analyzing their strengths, weaknesses, and future research trends. We then reviewed the existing benchmark data sets (e.g., ETH Dataset and Robotic 3D Scanning Repository) for TLS point cloud registration and summarized their limitations. Finally, a new benchmark data set was assembled from 11 different environments (i.e., subway station, high-speed railway platform, mountain, forest, park, campus, residence, riverbank, heritage building, underground excavation, and tunnel environments) with variations in the point density, clutter, and occlusion. In addition, we summarized future research trends in this area, including auxiliary data-guided registration, deep learning-based registration, and multi-temporal point cloud registration.

1. Introduction

Point clouds collected using terrestrial laser scanners (TLS) are important data for a variety of applications, such as tracking urban growth (Pieraccini et al., 2006; Vosselman and Mass, 2010; Gehrung et al., 2016), forest structure assessments (Liang et al., 2012; Kelbe et al., 2016; Polewski et al., 2018), landslide surveillance (Prokop and Panholzer, 2009; Huang et al., 2019), three-dimensional (3D) model reconstruction (Morgan and Habib, 2002; Jung et al., 2014; Oesau et al., 2014; Xu et al., 2017; Dong et al., 2018a), cultural heritage management (Montuori et al., 2014), and solar energy surveys (Redweik et al., 2013; Freitas et al., 2015; Huang et al., 2017). These applications require registration in order to transform point clouds with arbitrary coordinate systems into a common coordinate system that covers the entire scene completely (Xu et al., 2019). Precise co-registration of point clouds is also generally required for change detection tasks (Boerner et al., 2019; Hebel and Stilla, 2012; Tuttas et al., 2017).

Detecting geometric characteristics and identifying correspondences are two important tasks when registering TLS point clouds in arbitrary initial positions and orientations (Habib et al., 2010). However, the following important obstacles can hinder these two tasks. (1) Outliers and noise induced by moving artifacts (e.g., swaying trees, moving pedestrians, bicycles, and vehicles), missed points (e.g., gaps caused by occlusions), and variable point densities can greatly affect the precision of the derived spatial characteristics (e.g., key points, curves, planes, and patches) (Zai et al., 2017). (2) Overlapping point clouds collected from different viewpoints can result in variable point distributions at the same position, and thus the registration method must consider view-independent criteria to match features (Yu et al., 2015). (3) Limited overlapping between different scans will lead to insufficient common points between successive scans, so most of the predicted correspondences will be outliers (Cai et al., 2019). (4) Urban environments contain many repetitive structures (e.g., façades and windows) and homogeneous architectural layouts, thereby increasing the

* Corresponding authors.

E-mail addresses: bshyang@whu.edu.cn (B. Yang), hongchao.fan@ntnu.no (H. Fan).

ambiguity of possible matches (Theiler et al., 2015, Dong et al., 2018b). (5) Huge volumes of data containing billions of points will incur high computational costs that require highly efficient approaches (Theiler et al., 2014a, 2014b), and thus the method employed must be scalable, e.g., avoid pairwise matching of individual points (Yu et al., 2015). (6) Deep learning-based approaches are the standard for practical computer vision applications but there have been no major advances in 3D point cloud processing due to the lack of adequate large-scale benchmark data sets. The lack of benchmark data and consistent evaluation criteria are regarded as limiting factors that affect the objective analysis of different methods (Cheng et al., 2018).

Many studies have attempted to address these difficulties in order to increase the effectiveness, efficiency, and robustness of point cloud registration in the fields of photogrammetry and remote sensing, computer vision, and mobile robotics (Salvi et al., 2007; Tam et al., 2012; Pomerleau et al., 2015; Weinmann, 2016; Cheng et al., 2018). Previous reviews have considered 3D point cloud keypoint detectors and descriptors (Tombari et al., 2013; Guo et al., 2016). In particular, Salvi et al. (2007) surveyed the most common techniques for fusing 3D surfaces and verified the performance of these techniques in the presence of disturbances due to noise and outliers. Tam et al. (2012) also provided a comprehensive review of point clouds and mesh registration in terms of rigid or non-rigid registration. Pomerleau et al. (2015) reviewed the iterative closest point (ICP) variants used for fine registration during the last 20 years and their applications in mobile robotics. Cheng et al. (2018) presented a comprehensive review of multi-angle and multi-scale light detection and ranging (LiDAR) data registration in the fields of photogrammetry and remote sensing. However, none of these previous reviews focused specifically on TLS point cloud registration in the domains of photogrammetry, remote sensing, and computer vision.

Several benchmark data sets for TLS point cloud registration are publicly available, such as the TLS point clouds collected from five different scenes (i.e., office, arch, trees, courtyard, and façade scenes), ground truth data (i.e., pairwise transformation matrices), and source code provided by the Institute of Geodesy and Photogrammetry at ETH Zurich University (Theiler et al., 2014a, 2014b). Jacobs University Bremen also published their benchmark data sets for TLS point cloud registration, which were recorded using a RIEGL VZ-400 system on their campus, where 122 scans cover almost the whole campus¹. However, the existing benchmark data sets are limited in terms of the diversity of scenarios, data acquisition equipment employed, environmental organizations, and scene changes.

Our main findings are as follows.

- To the best of our knowledge, this is the first literature survey to focus on the registration of TLS point clouds. In contrast to previous reviews, we explored the state-of-the-art methods in the domains of photogrammetry, remote sensing, and computer vision.
- We developed a large-scale benchmark data set for TLS point cloud registration. The proposed benchmark data set comprises 115 point clouds and over 1740 million 3D points collected from 11 different environments (i.e., subway station, high-speed railway platform, mountain, forest, park, campus, residence, riverbank, heritage building, underground excavation, and tunnel environments) with variations in terms of the point density, clutter, and occlusion.
- In addition, we provide insights into future research trends, including large-scale benchmark data sets, auxiliary data-guided registration, deep learning-based registration, and multi-temporal point cloud registration.

The remainder of this paper is structured as follows. In Sections 2 and 3, we provide comprehensive surveys of the existing methods and

benchmark data sets for TLS point cloud registration. In Section 4, we describe the benchmark data set proposed for TLS point cloud registration in detail. In Section 5, we briefly discuss possible future research directions. Finally, we give our conclusions in Section 6.

2. Registration techniques for TLS point clouds

Several methods have been proposed to improve the robustness, effectiveness, and reliability of TLS point cloud registration, which can be divided into two groups comprising pairwise and multiview registration, depending on the amount of input point clouds (Dong et al., 2018b). Most pairwise and multiview point cloud registration methods employ a coarse-to-fine strategy (Guo et al., 2013). In particular, coarse registration algorithms are first applied to calculate the approximate six degrees of freedom orientation and the translation parameters between adjacent point clouds (Xu et al., 2019). Next, fine registration algorithms, such as the normal distribution transform (NDT) and its variants (Biber and Straßer, 2003; Takeuchi and Tsubouchi, 2006; Das and Waslander, 2012), and the ICP algorithm (Besl and McKay, 1992) and its variants (Yang et al., 2013; Pavlov et al., 2018; Tazir et al., 2018), are utilized to refine the approximate orientation and translation.

2.1. Pairwise coarse registration

2.1.1. Hand-crafted feature-based methods

The hand-crafted feature-based methods have two significant steps: extracting geometric characteristics (e.g., points, curves, planes, and surfaces) and identifying correspondences (Dong et al., 2018b). In step one, the geometric features (e.g., spin image (Johnson and Sing, 1999), fast point feature histograms (FPFHs) (Rusu et al., 2009), tri-spin-image (Guo et al., 2015), semantic feature lines (Yang et al., 2016), binary shape context (Dong et al., 2017), adaptive covariance (Zai et al., 2017), intersection lines (Stamos and Leordeanu, 2003), planes (Xiao et al., 2013), geometric primitives (Rabbani et al., 2007), curves (Yang and Zang, 2014), or patches (Ge and Wunderlich, 2016) are calculated as the primitives for registration. In step two, the corresponding geometric features are identified using various matching strategies (e.g., random sample consistency) (Fischler and Bolles, 1981), geometric consistency constraints (Tombari et al., 2010), search for inliers (Buch et al., 2014), non-cooperative game (Zai et al., 2017), and fast match pruning (FMP) (Cai et al., 2019). Finally, the rigid transformations between point clouds are calculated based on the corresponding geometric features.

For example, Stamos and Leordeanu (2003) first identified the intersection lines between adjacent planes and then estimated the transformation of the neighboring point clouds based on at least two corresponding intersection lines. Dold and Brenner (2006) calculated the transformation parameters based on more than three corresponding patches. Khoshelham et al. (2010) estimated the positions and orientations of different scans using a linear least-squares adjustment model for the corresponding planes. Rabbani et al. (2007) first extracted planes, spheres, and cylinders as the primitives, before aligning the scans by identifying the corresponding primitives. Weber et al. (2015) used the FPFH encodings of the geometric features of each keypoint and then utilized a polygon-based correspondence rejector to identify the corresponding keypoints. Yang et al. (2016) extracted the feature lines from buildings and pole-like artifacts, and matched them by combining their semantic and geometrical characteristics. Zai et al. (2017) generated covariance matrix descriptors for keypoint detection and descriptions to identify potential correspondences, before conducting a non-cooperative game to separate mutually compatible matches. Xu et al. (2017) utilized a geometric constraint constructed by building facades and ground surfaces for the coarse registration of urban point clouds. Cai et al. (2019) calculated the FPFH for each keypoint to identify potential correspondences and utilized FMP to reduce them significantly, while ensuring the preservation of the correct

¹ <http://kos.informatik.uni-osnabrueck.de/3Dscans/>

correspondences.

This type of approach has been used widely for pairwise coarse registration because the geometric characteristics have practical utility on different scenes (Dong et al., 2018b). However, this method still has the following common limitations. First, approaches based on feature lines, plane, and surfaces can obtain good performance in urban environments with man-made artifacts, but they lead to failure in wild landscapes (e.g., hill, woodlands, and river) with less structural characteristics. Second, the feature point-based approaches can achieve good performance in most the environments, but the methods for detecting and describing keypoints are vulnerable to changes in the point density and noise (Guo et al., 2014).

2.1.2. Four-points congruent set (4PCS)-based registration method

The 4PCS-based method (Aiger et al., 2008) allows the rapid and robust registration of frames by aligning 3D point sets using wide bases. Instead of using the randomly selected point sets employed by the RANSAC-based registration scheme, the 4PCS-based registration scheme utilizes particular point sets comprising congruent points of constant affine invariant ratios for the distances between pairs of points. In particular, 4PCS samples four points with an approximately co-planar layout as a base set, rather than the set of the minimal three points (Theiler et al., 2014a, 2014b). To find the corresponding sets in the source and target point clouds, 4PCS assumes that intersectional diagonal ratios are invariant under the affine transformation in the arbitrary planar quadrangle (Huttenlocher, 1991), and this constraint is applied to identify whether two base sets are matched or not. The 4PCS scheme can work well for data sets with small overlaps and it requires no assumptions regarding the initial positions. Many variants have been developed based on the original 4PCS, where the main improvements include reducing the number of candidate points (e.g., by simply using features points as candidates), enhancing the robustness (e.g., by adding more constraints than point-to-point distances), implementing large-scale scenes (e.g., by replacing points with higher-level structures in the congruent sets), and decreasing the failure rate (e.g., by relaxing the constraints when matching intersections).

For instance, Super4PCS (Mellado et al., 2014) is an improved more efficient variant of the original 4PCS that addresses two bottlenecks by decreasing the quadratic time complexity of 4PCS to linear time complexity in Super4PCS, and reducing the number of conjugate point pairs. The low time complexity of Super4PCS is achieved by applying a splitting and indexing strategy using a 3D grid and by limiting the search of point pairs to 3D spheres. The conjugate point pairs are reduced via angle constraints formed based on the directions of the point pairs mapped to a spherical shell map, which stores a set of vector directions. The main benefit of Super4PCS is its significantly accelerated runtime. Keypoint-based 4PCS (K-4PCS) (Theiler et al., 2014a, 2014b) is another variant of 4PCS where the efficiency and reliability are increased by reducing the number of unnecessary points. Instead of using arbitrary points as candidates or simply downsampling the data set, K-4PCS only detects and extracts discriminative points (i.e., keypoints) as input candidates for creating and finding corresponding congruent sets. The major advantage of K-4PCS is that it can significantly decrease the amount of candidates with a tractable computational cost because the number of keypoints is much smaller. However, the performance of K-4PCS is still highly dependent on the quality of the extracted keypoints and it is sensitive to variable point densities. Semantic-keypoint-based 4PCS (SK-4PCS) (Ge, 2017a, 2017b) is an improved version of K-4PCS that uses the semantics of keypoints as additional constraints. Thus, SK-4PCS detects feature points with semantic information (e.g., the corner of walls or connection of pillars) from different cross-sections of buildings. The complexity of the search for corresponding congruent sets is greatly decreased by introducing semantic information.

In contrast to the two variants mentioned above, generalized 4PCS (G-4PCS) (Mohamad et al., 2014) uses a different strategy to increase the reliability, where it relaxes the constraint on the co-planarity of the

four points in the congruent set. Thus, G-4PCS considers the non-co-planar four-points base set as planar points, before measuring and evaluating the distances between the intersections of point pairs instead of determining the coincidence of intersections. By relaxing this constraint, G-4PCS supports the fuzzy search of congruent sets to considerably enhance the reliability and robustness. The inverse relationship between the intersection length magnitude and G-4PCS sets has been demonstrated experimentally. Furthermore, by combining the advantages of Super4PCS and G-4PCS, Mohamad et al. (2015) proposed super-generalized 4PCS, which significantly speeds up the process compared with Super4PCS. In addition, to further reduce the number of candidates, structural elements such as planes can be introduced to generate the congruent sets. Four-planes congruent sets (Bueno et al., 2018) and voxel-based four-planes congruent sets (Xu et al., 2019) are representative methods that use planes instead of points as the basic elements to form congruent sets. In order to obtain the plane-based congruent set, the ratios of the plane pairs are calculated based on the angles between the planes and intersecting planes. The fitting or extraction of planes from the points can avoid uneven point densities and outliers, thereby ensuring the reliable orientation of the planes and angle measures. In contrast, the distances between the points used by traditional 4PCS methods are more sensitive to outliers and noise in the data sets. Similarly, volumetric 4PCS (V-4PCS) (Huang et al., 2017) extends the 4PCS algorithm from four coplanar points to non-coplanar volumetric information. The computational complexity is significantly reduced with V-4PCS because the number of pre-clustered planes is much less than that of points.

The 4PCS-based methods have two main weaknesses. First, the 4PCS scheme is actually a strategy for reducing the candidate sets by retrieving invariant congruent subsets, where it relies on a robust estimator to match the correspondences in order to reject mismatches and outliers, which is a time-consuming procedure. Second, it is assumed that the congruent sets are invariant to the rigid transformation. The congruent sets can maintain constant ratios (e.g., between the distances or angles) and provide identical intersections (e.g., point or vector) for matching the congruent sets only under this prior condition that requires rigid transformation. The criteria for matching the intersections of congruent sets should be relaxed in applications with non-rigid transformations (Ge, 2017a, 2017b). Indeed, experimental investigations have shown that the 4PCS-based methods can be improved by modifying the definitions of congruent sets, which should consider relaxed criteria in non-rigid transformation situations. Essential feature descriptions (e.g., the geometry of keypoints) for the elements of congruent sets should also be developed.

2.1.3. Probabilistic registration methods

In contrast to the hand-crafted feature-based methods, probability registration methods employ relaxation matching to identify correspondences. In this approach, one-to-many matching patterns are allowed so more correspondence mapping combinations can be considered, thereby improving the flexibility and robustness of the alignments. In general, probabilistic registration methods model the distribution of the point clouds as a density function, before optimizing the density function with a correlation-based approach or expectation maximization-based (EM) optimization framework (Lawin et al., 2018).

Tsin and Takeo (2004) proposed an affinity measure called kernel correlation for point cloud registration in correlation-based approaches. First, the kernel functions are utilized to model the distributions of point clouds. A cost function describing the similarity between them is then constructed and optimized by using the kernel correlation as an M-estimator. This method employs a built-in smoothing algorithm, which makes it robust to noise and outliers. Experiments have demonstrated its robustness and resistance to noise. However, the selection of the kernel function is important and the method employed still needs to be improved. Jian and Vemuri (2010) used Gaussian mixture models (GMMs) to represent the distributions of point clouds. The registration

problem was then transformed into aligning two GMMs by minimizing the statistical discrepancy between them. To register the GMMs, they employed the L2 distance and a robust estimator to minimize the L2 distance between GMMs. This approach is a closed-form expression and easy to implement, thereby yielding a computationally efficient registration algorithm.

In the EM-based approaches, Myronenko et al. (2007) and Myronenko and Song (2010) introduced a coherent point drift (CPD)-based method for point cloud registration, which considers the registration problem as a probability density estimation problem. CPD-based methods use GMM centroids to describe the source point cloud and fit the GMM to the target point cloud by maximizing the likelihood of the objective function. In this algorithm, it is assumed that the GMM centroids move coherently in order to preserve the geometric structure of the point cloud. Many mathematical tricks (e.g., likelihood transformation and bound estimation) are used to simplify the objective function. A closed-form solution of EM algorithm is derived to solve the matching matrix and the transformation parameters in a successive manner. CPD exhibits generality, accuracy, and good robustness to noise and outliers. The variants of CPD mainly consider the automatic estimation of the percentage of outliers (Wang et al., 2011), refinement of correspondences by geometric constraints (Zang and Lindenbergh, 2019), and improving the computational efficiency (Lu et al., 2014). Within this probabilistic registration framework, Golyanik et al. (2016) presented the extended CPD (ECPD) registration algorithm to broaden its scope. They constructed a novel closed-form model to couple the correspondence priors into the registration procedures. ECPD uses the correspondences while maintaining subsampling by counterbalancing the polynomial complexity and reducing the number of operations.

However, the existing probabilistic registration methods cannot deal with large volume points simultaneously because the registration result depends on the sampling result. Thus, an incremental registration method that allows the Gaussian distributions to grow should be developed in future research. In addition, the estimation of the outlier ratio affects the accuracy and efficiency of registration, but few methods can automatically estimate the outlier ratio. Probabilistic registration methods will be applied much more widely if a more effective algorithm for outlier ratio estimation can be developed.

2.1.4. Deep learning-based registration methods

The traditional methods use handcrafted features to distinguish correspondences, and thus they are affected by the experience of their designers and the capacity for parameter adjustment. Therefore, the generalizability, robustness, and descriptive capacity of handcrafted features may not be sufficient. Deep learning-based methods can directly learn high-level feature representations from a large volume of data to achieve good performance in terms of both the descriptive capacity and robustness to variations in the point density and viewpoints.

Deep learning-based registration methods can be divided into three categories according to their data representations comprising voxelization-based methods, multiview-based methods, and irregular point-based methods. The voxelization-based registration methods first divide point clouds into regular 3D grids, before utilizing the 3D convolution network structure to calculate the features and determining the corresponding points with a predefined feature similarity metric (Zeng et al., 2017; Zhang et al., 2019). These methods achieve good results with small-scale indoor data sets. However, voxelization-based methods lead to information loss, as well as consuming large volumes of memory and computational resources, and their accuracy is affected by the resolution of the 3D grid. Multiview-based methods are inspired by two-dimensional (2D) image registration methods, where they first represent the point cloud as a set of images rendered from different views, before applying image convolutional neural networks to calculate the features and determine the correspondences based on the learned features (Zhou et al., 2018; Pujol-Miró et al., 2019). These methods depend greatly on the projection directions, neighborhood definition, and

projected image resolution, where some important information is lost in the projection process, so the generalizability is limited.

PointNet (Qi et al., 2017a) and PointNet++ (Qi et al., 2017b) are pioneering methods for directly processing unordered point sets, which are invariant to transformations by T-Net and to different permutations of the data feeding order using a symmetry function. Based on PointNet (Qi et al., 2017a), the PPFNet method (Deng et al., 2018a) first constructs patches by combining the coordinates, normal vectors, point pair features, and neighboring points, before PointNet is utilized to calculate the features of each patch and determine the corresponding relationship between each patch with a random sample consensus algorithm (Fischler and Bolles, 1981). PPF-FoldNet (Deng et al., 2018b) also uses the point pair features as inputs, as well as mini-PointNet and skip-links (He et al., 2016) as the encoder, and FoldingNet (Yang et al., 2018) as the decoder to reconstruct the point PPFs. PointNetLK (Aoki et al., 2019) introduces the Lucas-Kanade algorithm into 3D point cloud registration and solves the problem iteratively with PointNet. However, this method involves iterative computations that depend on the thresholds of the optimal twist parameters. Deng et al. (2019) proposed an end-to-end architecture for solving the problem of registering two point cloud scans by employing FoldingNet (Yang et al., 2018) for feature encoding and multi-layer perceptron (MLP) for estimating the transformation. The deep closest point method (Wang and Solomon, 2019) employs a dynamic graph convolutional neural network (Wang and Solomon, 2019) for feature extraction and an attention module to generate a new embedding that considers the relationships between two point clouds. In addition, a singular value decomposition module (Papadopoulou and Lourakis, 2000) is used to calculate the rotation and translation.

These methods achieve good performance at small-scale indoor point cloud registration, but they are limited by the number of samples and their capacity to extract features from the model, and they are also difficult to generalize to outdoor data. Only a few deep learning models are suitable for the registration of outdoor scenes. For example, 3DFeatNet (Yew and Lee, 2018) is a weakly supervised deep learning method for registering point clouds in the KITTI data set (Geiger et al., 2012), where the point cloud is first divided into anchor points, positive points, and negative points, before applying detector (i.e., multi-layer perceptron and max-pooling) and descriptor networks to determine the correspondences in the anchor and positive points. DeepICP (Lu et al., 2019) is an end-to-end architecture for registering two point clouds in the KITTI data set, which is based on the PointNet, PointNet++, and Voxnet methods. However, this method is limited by the initial pose accuracy of the two point clouds.

The deep learning method has proved effective for the registration of indoor and small-scale outdoor point clouds. However, limitations on the amount of data and complexity mean that it is difficult to apply to large-scale outdoor point clouds. Compared with methods based on handcrafted features, the deep learning-based methods can automatically learn more robust and distinct feature representations, and thus they have great potential for registering scenarios containing extremely repetitive and symmetric elements, weak geometric features, and limited overlaps.

2.2. Pair-wise fine registration

2.2.1. ICP and its variants

The ICP algorithm (Besl and McKay, 1992) is the most commonly used method for pairwise fine registration because of its conceptual simplicity and high usability. Due to its good initial transformation, ICP performs locally optimal registration by conversely resolving the nearest point-to-point correspondence and optimal rigid transformation until convergence. However, its application to LiDAR point clouds is challenging because there are no exact one-to-one correspondences between two point clouds (Cheng et al., 2018). In addition, these point clouds often contain a huge number of points and they are corrupted by

variations in the point density, noise, and outliers due to partial overlapping, occlusion, and clutter. Therefore, numerous variants of ICP have been proposed to address these challenges.

To improve the accuracy of ICP during registration, variants have been proposed using **point-to-plane** (Chen and Medioni, 1991), **point-to-projection** (Campbell and Flynn, 2001), and **plane-to-plane** (Segal et al., 2009) correspondence error metrics. To determine the correspondences more efficiently and accurately, Bae and Lichten (2008) proposed a geometric primitive ICP with random sample consensus by using the curvature and normal vector for correspondence matching. Gressin et al. (2013) used the geometric features of the neighborhood to improve the accuracy and speed of ICP. Bouaziz et al. (2013) proposed sparse ICP where the ICP problem is expressed as a sparse l_p metric optimization problem, and they obtained a robust and heuristic-free registration algorithm that systematically addresses the problem of outliers.

Some studies focused on broadening the basin of convergence and improving the robustness to outliers, low overlap, and occlusions. Thus, Chetverikov et al. (2002) proposed a **trimmed ICP (TrICP)** algorithm, which uses the least trimmed squares approach to sort the square errors and minimize a certain number of smaller values. This method considers the outliers, shape defects, or partial overlaps, and the optimal transformation is estimated at each iteration step, thereby making it highly robust to incomplete and noisy data. Sharp et al. (2002) proposed an ICP method with invariant features by combining invariant features with the geometric distance in the closest distance calculation, and it is more likely to converge to the global optimum than ICP under ideal, noise-free conditions. Fitzgibbon (2003) optimized the ICP errors by using the nonlinear optimization Levenberg-Marquardt ICP algorithm, which enlarges the range of convergence and avoids any obvious loss of speed. Yang et al. (2013) introduced the **global optimal ICP method (Go-ICP)** to integrate ICP with a branch-and-bound scheme, so coarse registration is not needed. However, Go-ICP is much more time consuming than ICP and it is sensitive to outliers. Dong et al. (2014) introduced the LieTrICP algorithm, which combines the Lie group with the TrICP method. After determining the correspondences, LieTrICP uses the Lie group to minimize the trimmed squared distances of the point pairs in order to estimate the transformation. The application of the Lie group allows the algorithm to deal with the data by anisotropic scale transformation and with a low overlap rate.

Another class of ICP variants is focused on improving the efficiency. Thus, Simon et al. (1995) proposed an ICP algorithm based on decoupled acceleration. This algorithm refines the near optimal data and selects the lowest amount of data for pose refinement to preserve the accuracy. Qiu et al. (2009) proposed an accelerated ICP method based on GPU units, which involves constructing an array-based k-d tree, and using a priority search method and single-element priority queue to implement the GPU calculation, thereby ensuring that real-time calculations can be performed. Uhlenbrock et al. (2017) used a 2D array cache for storing local subtrees during k-d tree search to speed up the iterative registration process. This ICP caching method is computationally efficient for local subtree caching. Tazir et al. (2018) proposed a novel approach for sparse to dense point cloud registration by using a single point in one cluster as the keypoint. Pavlov et al. (2018) proposed the use of the Anderson acceleration technique in ICP, which speeds up the convergence process in ICP by finding a fixed point in the contractive mapping that meets the desired quality. This method requires less iterations to converge to the same error compared with simple iteration.

The two main limitations of ICP are as follows. First, ICP is highly dependent on a good initialization step in order to prevent the algorithm becoming trapped by a local optimum. Second, incorrect closest-point correspondences are very common in the registration process due to the local optimal correspondence matching strategy, thereby limiting the accuracy of registration. To the best of our knowledge, no variant of ICP can cope with these two limitations simultaneously.

2.2.2. NDT and its variants

The NDT algorithm is an alternative approach for the fine registration of point clouds. The NDT algorithm was first used for the localization and mapping of robotics in 2D space (Biber and Straßer, 2003), before it was extended to 3D space (Magnusson and Duckett, 2005; Takeuchi and Tsubouchi, 2006). In contrast to the ICP algorithm, the referenced point clouds are represented by a set of Gaussian distributions with different probability density functions (PDFs) in the NDT algorithm. According to the desired transformation between the reference and the source point clouds, a certain cost is obtained that evaluates the alignment quality between the two point clouds. The cost is determined based on the similarity between the PDFs of the reference point clouds and transformed point clouds. Using the NDT algorithm, the original fine registration problem is transformed into a nonlinear optimization problem, where it determines the optimal transformation by maximizing the similarity between the PDFs. In particular, the similarity between PDFs can be evaluated in a point-to-distribution manner (Magnusson et al., 2007) or distribution-to-distribution manner (Stoyanov et al., 2012). The computational time and convergence basin are both improved by conducting the evaluation in a distribution-to-distribution manner (Stoyanov et al., 2012). A key step in the NDT algorithm is establishing the grids with PDFs because the grid size is difficult to determine. Thus, a method was proposed for registering the point clouds using NDT at multiple scales (Takeuchi and Tsubouchi, 2006; Magnusson et al., 2007; Das and Waslander, 2012). The multi-scale NDT is an effective method that converges to the optimal solution and it can obtain high quality results.

The ICP and its variants require a high point density to obtain accurate correspondences from the nearest neighbor (Nüchter, 2008), so the NDT algorithm is more reasonable for registering point clouds with low point densities because the sparsity challenge can be partly solved by modeling the point clouds as Gaussian distributions. However, the NDT has a shortcoming because its cost function is discontinuous. The NDT represents the point clouds as grid cells with PDFs and a transformed point may cross cells during nonlinear optimization when the transformation is updated. The following methods have been proposed to handle this discontinuity problem. Biber and Straßer (2003) used overlapping grid cells and Magnusson (2009) interpolated the PDF between grid cells using trilinear interpolation. The overlapping grid cells and interpolation method improve the original NDT algorithm, but they still cannot fully solve the discontinuity problem. Thus, Das and Waslander (2014) utilized the NDT based on segments instead of grid cells to solve the discontinuity problem. However, the segmentation of point clouds is time consuming and it introduces uncertainty into the registration results.

The main limitations of the NDT algorithm are similar to those of the ICP algorithm. First, the NDT algorithm is also highly dependent on an accurate initial value. Second, a transformed point may cross PDF cells during nonlinear optimization when the estimated transformation is updated, and the transformed point may be located in the wrong corresponding PDF cell during this process.

2.3. Multiview point cloud registration

The overlaps and orders of point clouds are unknown during multiview point cloud registration, and this problem is significantly more difficult than pairwise registration, although it has received relatively little attention (Dong et al., 2018b). The multiview registration methods involve three difficult challenges, i.e., attempting to recover the view orders, estimating pairwise rigid transformations, and determining absolute poses, and they can be grouped into sequential registration-based methods and joint registration-based methods (Min and Meng, 2019).

2.3.1. Sequential pairwise registration-based methods

The sequential pairwise registration-based methods conduct

register-then-integrate processes until the point clouds are all translated into a common system (Ge et al., 2019; Min and Meng, 2019), and they can be further divided into minimum spanning tree-based (MST) methods, shape growing-based methods, hierarchical merging-based methods, and motion averaging-based methods.

Many studies have focused on MST multiview registration methods (Huber and Hebert, 2003; Zhu et al., 2016; Kelbe et al., 2016; Cai et al., 2019). For example, in the methods proposed by Weber et al. (2015) and Yang et al. (2016), the FPFH and semantic feature point descriptors are first calculated for exhaustive pairwise coarse alignment to create a weighted graph with fully connected edges, where the overlaps or number of correspondences are treated as the edge weight. The MST (Kruskal, 1956) is then constructed for the graph to determine the registration paths from each point cloud to the anchor. Finally, the input point clouds are registered in a common coordinate system based on the pairwise transformations and registration paths. These types of algorithms avoid conflicts by constructing the MST that connects each scene to the anchor, but they do not utilize the duplicate information supplied by multiple overlaps or loops to minimize the registration error. To address potentially inconsistent matches, Kelbe et al. (2016) proposed the use of the Dijkstra distance to build MSTs and their registration results are combined. In particular, their method employs an embedded confidence metric to evaluate the pairwise registration results, before constructing the weighted graph and its Dijkstra spanning trees based on the pairwise transformations. Finally, the competing Dijkstra spanning trees are arranged in a common system of coordinates. However, the MST-based methods begin with a time-consuming exhaustive pairwise registration process, which is unsuitable for large-scale TLS point cloud registration (Ge et al., 2019).

Several shape growing-based algorithms have been introduced to improve the efficiency of MST-based methods (Zhu et al., 2016). For instance, the method proposed by Mian et al. (2006) first selects the point cloud as the root node with the largest surface area, before sequentially integrating other point clouds with sufficient correspondences. Using a similar procedure, the method described by Guo et al. (2014) first selects a seed scan and then incrementally merges the scans by conducting pairwise registration between the seed and the remaining scans until all of the scans are transformed into a common coordinate system. In general, the shape growing-based methods are more efficient compared with MST-based methods. Ge and Hu (2020) proposed an object-based incremental registration strategy for TLS point clouds in urban environments that increases the reliability of the 2D transformation step using line primitives and it achieves the global optimum by least-squares optimization.

Several hierarchical merging-based methods have been proposed to enhance the robustness and efficiency with limited overlaps (Tang and Feng, 2015; Zhu et al., 2017; Dong et al., 2018b). For instance, Tang and Feng (2015) solved the problem of multiview registration by conducting hierarchical optimization based on an undirected graph. In particular, the proposed method first aligns each overlapped point cloud pair locally, before hierarchically refining them with an incremental and loop-based registration algorithm, and then redistributing the accumulated errors to all views by global error diffusion. Zhu et al. (2017) introduced a hierarchical multiview registration method based on the overlapping percentages in a spanning tree. The root scan and its connected scans are then recursively registered, and merged by updating the spanning tree. Dong et al. (2018b) developed a hierarchical merging-based method to align multiview and unordered point clouds without prior knowledge of their positions and orientations. The proposed method significantly reduces the computational complexity by employing a global vector of the locally aggregated descriptor (Jegou et al., 2012) to assess the similarity of the two scans, and it also improves the capacity to handle point clouds with small overlaps by exploiting the overlaps between multiple point clouds.

The sequential pairwise registration-based methods are problematic because errors accumulate in the individual steps, thereby leading to

poor predictions as the model grows in size (Evangelidis and Horaud, 2017). Several joint registration-based methods have been developed to overcome these limitations.

2.3.2. Joint registration-based methods

The multiview registration problem is treated as a graph optimization issue where each node and each edge denote a point cloud and a pairwise registration between adjacent point clouds, respectively (Ge et al., 2019). Based on graph theory, Boykov et al. (2001) introduced a global energy optimization method where all edges and loop closures are considered to improve the registration accuracy. In the method proposed by Theiler et al. (2015) using the exhaustive pairwise registration graph, an energy function is first constructed by considering all of the loop consistency constraints in the graph, before the pair transformations are disambiguated using the lazy flipper algorithm to minimize the energy function (Andres et al., 2012). Both of these methods only redistribute the registration error over the graph built without updating the correspondences and they cannot decrease the total registration errors.

Given consistent and accurate relative motions, the motion average algorithm can achieve accurate multiview registration using the Lie algebra motions framework. Shih et al. (2008) treated multiview registration as a quadratic programming problem with Lie algebra by distributing the accumulated error using the constraints on every closure loop based on a quadratic model. Govindu and Pooja (2013) simultaneously solved the problems of identifying correspondences and estimating their 3D motions by distributing the errors evenly across the scans using Lie group structures of motions. Arrigoni et al. (2016) treated the multiview registration problem as a low rank sparse matrix decomposition problem. In particular, their proposed method first estimates the relative motions for each scan pair with large overlaps and ignores other motions, before the ignored motions are recovered by low rank sparse decomposition. This method performs well in the presence of outliers, but it is highly vulnerable to the sparseness of the incomplete matrix (Zhu et al., 2019).

The main disadvantage of motion averaging-based methods is that the pairwise registration process may contain outliers and noise, which can have negative effects on the estimation of the registration parameters (Evangelidis and Horaud, 2017). Recently, Evangelidis and Horaud (2017) suggested a clustering-based method for the joint registration of multiple point clouds called JRMP. JRMP assumes that all input points are extracted from a central Gaussian mixture, thereby transforming registration into a soft clustering problem. A batch and incremental EM algorithm was developed to calculate the parameters of the GMM and the transformations that align the points optimally. Theoretically, this method has good potential for reconstructing surfaces from multiview point clouds. Inspired by JRMP, Zhu et al. (2019) treated multiview registration as a clustering-based problem, which their method solves using a modified K-means algorithm. Alternate and iterative clustering and transformation calculations are applied to improve the efficiency and robustness. Min and Meng (2019) proposed a novel probabilistic approach for jointly registering multiple point clouds to enhance the robustness to noise and outliers. In particular, the proposed method first employs the orientation information (e.g., normal vectors) and position information associated with each point as inputs, before applying a hybrid mixture model obtained by combining Gaussian and von Mises-Fisher distributions to model the position and orientation of the multiple point clouds, and finally solving a HMMs problem under the EM framework to jointly register the multiple point clouds.

Theoretically, the most accurate multiview registration results can be obtained using joint registration-based methods. However, these methods are readily trapped by local minima due to the large calculation and search spaces.

3. Benchmark for TLS point cloud registration

Benchmarking efforts have a long tradition in the photogrammetry, remote sensing, and computer vision community. Due to the rapid development of deep learning-based methods, large-scale point cloud benchmark data sets for semantic segmentation (e.g., Oakland 3-D,² S3DIS,³ and Semantic 3D⁴), instance segmentation (e.g., Paris-Lille-3D⁵ and SynthCity⁶), and object detection (e.g., KITTI Vision Benchmark⁷) containing billions of points have become the standards for deep learning-based point cloud processes. These publicly available data sets have greatly enhanced research into point cloud classification and segmentation.

Only a few public data sets are available for TLS registration (e.g., ETH PRS TLS data set⁸), but some other public data sets can also be used for testing TLS registration methods, such as the ETH ASL Datasets Repository,⁹ Robotic 3D Scan Repository,¹⁰ and UNAVCO TLS Archive.¹¹ However, these data sets are highly heterogeneous and limited in terms of their data volume, environmental diversity, and collection devices. In the following, we provide detailed introductions to the existing data sets as well as identifying their strengths and weaknesses for TLS point cloud registration. Brief summaries of these data sets are presented in Table 1.

3.1. ETH PRS TLS data set

This data set was published for testing TLS registration methods, with 32 scans from five different scenes scanned by Z + F Imager 5006i and Faro Focus 3D. The limited dimensions lead to relatively high overlap rates between the scans. In addition, there are a few repetitive and symmetric elements, and moving objects. These limitations make the data set less challenging.

3.2. ETH ASL datasets Repository

This data set was collected to verify registration algorithms for point clouds obtained in specific environments and conditions. The data set contains eight sequences each with around 35 point clouds. The different sequences are characterized by diverse environments and geometric primitives. However, the data were collected using a custom-made rotating scanner (Hokuyo UTM-30LX) with a theodolite, so the accuracy and point density are lower compared with normal TLS point clouds.

3.3. Robotic 3D scan repository

This data set was mainly produced for developing simultaneous localization and mapping (SLAM) or interpretation algorithms. The 14 data sets scanned by Riegl VZ-400 were manually registered using markers, so they can be used for registration. However, although more than 200 scans are provided, these scans were mainly conducted in a structured environment with many artificial elements.

3.4. UNAVCO TLS Archive

This data set has been maintained by UNAVCO since 2008 in order to provide TLS data for geodesy-based research. The archive provides access to both raw (L0) TLS data and merged, aligned, and georeferenced (L2) point cloud data, so these data are also suitable for testing TLS registration methods. Data from more than 100 projects have been published and the number is updated continually. However, most of the scans were conducted in unstructured environments with natural elements.

4. Large-scale TLS point cloud registration benchmark: Wuhan University TLS (Whu-TLS) data set

The proposed benchmark called the Whu-TLS data set comprises 115 scans and over 1740 million 3D points collected from 11 different environments (i.e., subway station, high-speed railway platform, mountain, forest, park, campus, residence, riverbank, heritage building, underground excavation, and tunnel environments) with variations in the point density, clutter, and occlusion. The aims of the proposed benchmark are to facilitate better comparisons and provide insights into the strengths and weaknesses of different registration approaches based on a common standard. The ground-truth transformations, transformations calculated by Dong et al. (2018b), and registration graphs are also provided to allow researchers to evaluate their registration solutions and for environmental modeling. In addition, the Whu-TLS data set provides suitable data for applications in safe railway operation, river surveys and regulation, forest structure assessment, cultural heritage conservation, landslide monitoring, and underground asset management. We hope that the Whu-TLS data set meets the needs of the research community and becomes an important data set for the development of cutting-edge TLS point cloud registration methods. Table 2 shows detailed descriptions of the Whu-TLS data set in terms of the data acquisition equipment, number of scans, number of points, range covered, organization of the environment, location of the environment, and the minimum and maximum overlaps between pairwise point clouds. According to Pomerleau et al. (2014), the organization of the environment can be characterized as structured, unstructured and semi-structured, and the locations of the environment can be outdoors and indoors. Figs. 1–11 illustrates the sampled point cloud and image in each scene, and the globally consistent point clouds for the whole scene. The Whu-TLS data sets are available at <http://3s.whu.edu.cn/ybs/en/benchmark.htm>.

It should be noted that the Whu-TLS data set is challenging because: (1) multiple laser scanner systems (i.e., VZ-400, IMAGER5010C, ScanStationC5, LeicaP40, and Leica HDS6100) with differences in terms of the measurement range, accuracy, and field of view were used to collect the data; (2) it covers a wide range of scenes (i.e., subway station, high-speed railway platform, mountain, forest, park, campus, residence, river bank, heritage building, underground excavation, and tunnel environments) with significant difference in the organization of the environment and geometric shapes; (3) the overlaps between adjacent point clouds are extremely low (e.g., the minimum overlaps for the tunnel data set and residence data set are 5.5% and 1.0%, respectively); (4) some data sets include extremely repetitive and symmetric elements (e.g., tiles on the wall in the tunnel data set and railway tracks in the high-speed railway data set), as well as weak geometric features (e.g., the roof of the tunnel), which add uncertainty to the correspondence matching process; and (5) the changes in the scenes caused by moving objects (e.g., campus and heritage building data sets) and seasonal changes (e.g., mountain data set) also challenge the robustness of registration algorithms.

4.1. Whu-TLS subway station data set

The Whu-TLS subway station data set was collected in a subway

² https://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/

³ <http://buildingparser.stanford.edu/dataset.html>

⁴ <http://www.semantic3d.net/>

⁵ <http://npm3d.fr/paris-lille-3d>

⁶ <http://www.synthcity.xyz/>

⁷ http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d

⁸ https://prs.ipg.ethz.ch/research/completed_projects/automatic_registration_of_point_clouds.html

⁹ <https://projects.asl.ethz.ch/datasets/>

¹⁰ <http://kos.informatik.uni-osnabrueck.de/3Dscans/>

¹¹ <https://tls.unavco.org/>

Table 1

Brief summaries of existing benchmark data sets.

| Dataset | #scenes | Scanner | Objective | Character |
|-----------------------------|---------|----------------------------------|------------------|-------------------------------|
| ETH PRS TLS Dataset | 5 | Z + F Imager 5006i Faro Focus 3D | TLS registration | Limited amount and dimensions |
| ETH ASL Datasets Repository | 8 | Hokuyo UTM-30LX | Registration | Not TLS |
| Robotic 3D Scan Repository | 14 | Riegl VZ-400 | SLAM | Most in structured |
| UNAVCO TLS Archive | >100 | Riegl VZ series et al. | Geodesy research | Most in unstructured |

Table 2

Details of the Whu-TLS benchmark data set.

| Name | Scanner | #Scans | #Pts (million) | Dimensions (m) | Texture | Environment organization | Environment locations | Overlap (%) | |
|----------------------------------|----------------|--------|----------------|---------------------------------|---------|--------------------------|-----------------------|-------------|------|
| | | | | | | | | Min | Max |
| Whu-TLS Subway station | IMAGER 5010C | 6 | 237.57 | X:557.88, Y: 270.138, Z: 15.67 | ✓ | Structured | Indoors | 23.7 | 64.3 |
| Whu-TLS Railway | VZ-400 | 8 | 49.86 | X:1096.92, Y:1193.28, Z: 166.10 | ✓ | Structured | Outdoors | 10.9 | 66.1 |
| Whu-TLS Mountain | ScanStation C5 | 6 | 19.61 | X:350.30, Y:340.18, Z: 162.73 | ✗ | Unstructured | Outdoors | 13.4 | 42.3 |
| FGI-TLS forest | Leica HDS6100 | 5 | 149.45 | X:86.83, Y:85.74, Z: 25.56 | ✗ | Unstructured | Outdoors | 34.6 | 55.5 |
| Whu-TLS Park | VZ-400 | 32 | 160.24 | X:897.55, Y:834.79, Z: 166.44 | ✗ | Semi-structured | Outdoors | 24.4 | 82.8 |
| Whu-TLS Campus | VZ-400 | 10 | 109.05 | X:799.82, Y:861.56, Z:218.73 | ✗ | Semi-structured | Outdoors | 5.6 | 49.6 |
| Whu-TLS Residence | Leica P40 | 7 | 43.70 | X:481.69, Y:500.11, Z:101.61 | ✓ | Semi-structured | Outdoors | 1.0 | 91.4 |
| Whu-TLS River bank | VZ-400 | 13 | 93.11 | X:695.07, Y:554.05, Z:165.94 | ✗ | Unstructured | Outdoors | 22.6 | 49.6 |
| Whu-TLS <i>Heritage building</i> | VZ-400 | 9 | 238.16 | X:790.99, Y:541.35, Z:163.49 | ✓ | Structured | Outdoors | 28.7 | 69.4 |
| Whu-TLS Excavation | VZ-400 | 12 | 482.42 | X:269.52, Y:308.76, Z:47.74 | ✗ | Unstructured | Outdoors | 9.0 | 72.8 |
| Whu-TLS Tunnel | VZ-400 | 7 | 157.02 | X: 483.2, Y: 307.9, Z: 75.7 | ✗ | Unstructured | Outdoors | 5.50 | 32.0 |

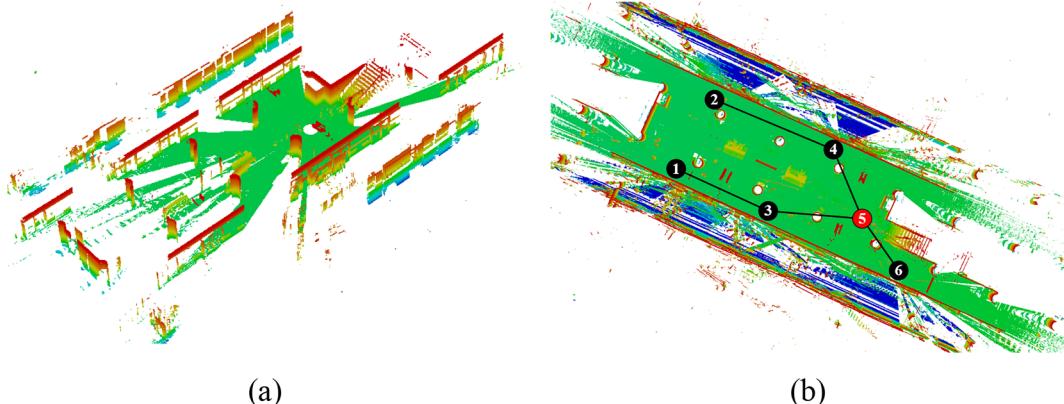


Fig. 1. Whu-TLS subway station data set: (a) sampled point cloud with shading reflecting differences in height (the roof of the subway station was deleted to improve the visualization); and (b) overview of the subway station and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

station in Shanghai using the Z + F IMAGER® 5010C laser scanner system, which integrates a camera and laser scanner to combine brilliant colors with high-resolution scanning data. This data set contains six point clouds and 237.57 million points, with a minimum overlap of 23.7% and maximum overlap of 64.3%. The environment mainly contains structural elements (e.g., elevators, pillars, and walls) and moving people, thereby providing a good example where registration algorithms must overcome the challenges due to the presence of both symmetric structures and moving objects. In addition, this data set contains a massive amount of virtual points caused by mirror reflections, which further increase the difficulty of registration. Fig. 1 illustrates a sampled point cloud, an overview of the whole scene, the position of each scan, and the registration path.

4.2. Whu-TLS high-speed railway data set

The Whu-TLS high-speed railway data set containing eight scans

and 49.86 million points was collected using the RIEGL VZ-400 laser scanner system. The environment mainly contains structural elements (e.g., railway tracks, utility poles, and platforms) and unstructured elements (e.g., trees and rocks). This data set provides a challenging environment containing symmetric structures (i.e., railway tracks). It also provides a suitable data set for the reconstruction of 3D models of railway tracks and safe railway operation applications. Fig. 2 illustrates a sampled point cloud, an overview of the whole scene, and the registration graph.

4.3. Whu-TLS mountain data set

Leica ScanStation C5 was used to collect the Whu-TLS mountain data set, which contains six point clouds and 19.61 million points, with a minimum overlap of 13.4% and maximum overlap of 42.3%. There are no significant vertical structures in the area, where unstructured rocks and grass are the main features. The point clouds were collected

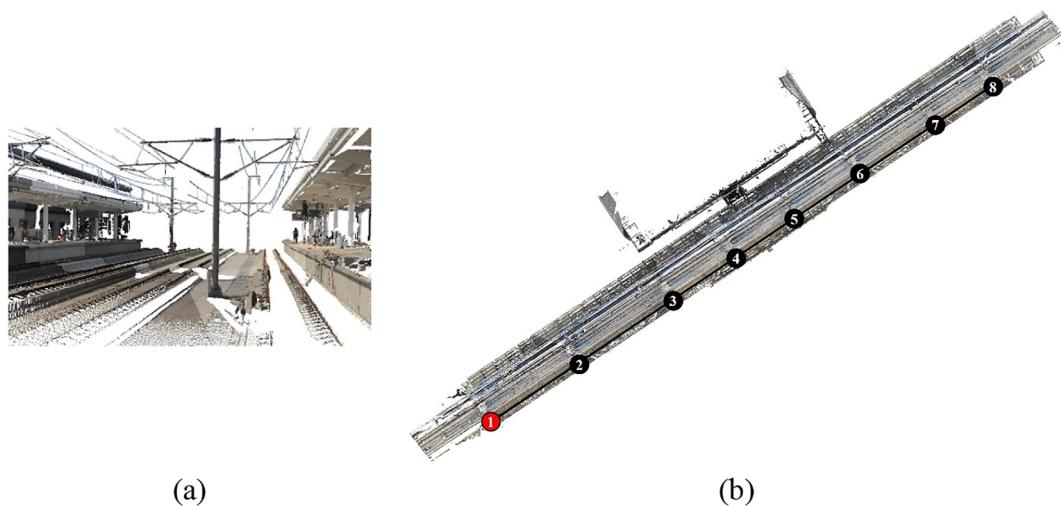


Fig. 2. Whu-TLS high-speed railway data set: (a) sampled point cloud; and (b) overview of the high-speed railway data set and registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

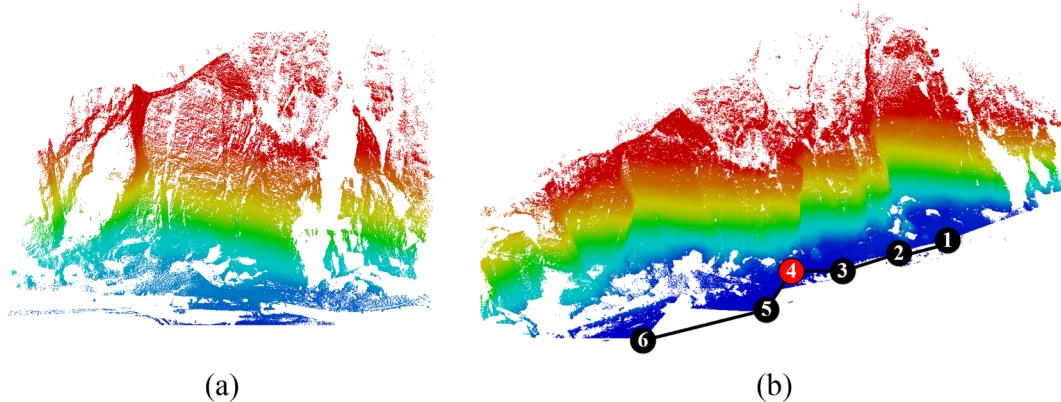


Fig. 3. Whu-TLS mountain data set: (a) sampled point cloud with shading reflecting differences in height; and (b) overview of the mountain data set and registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

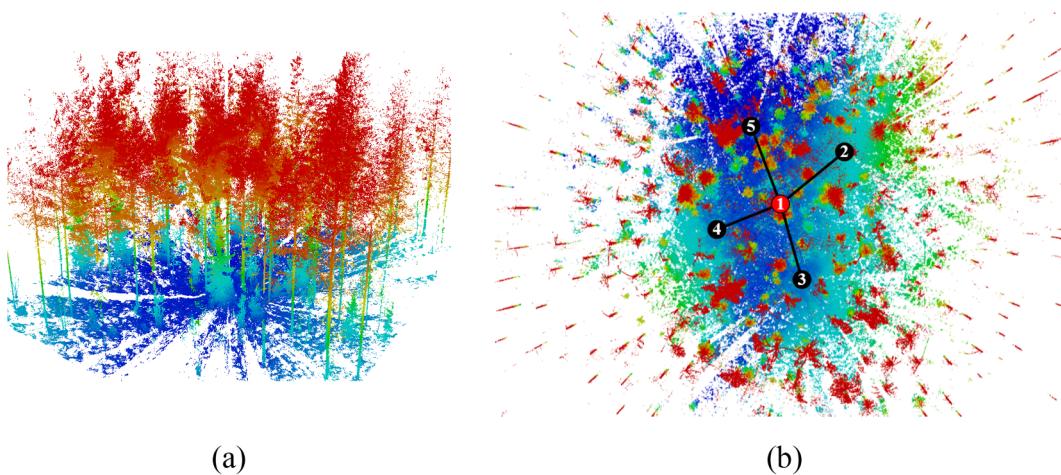


Fig. 4. FGI-TLS forest data set: (a) sampled point cloud with shading reflecting differences in height; and (b) overview of the forest data set and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in different seasons (i.e., point clouds 1–4 were collected in March and point clouds 5 and 6 were collected in August) and the changes in the details considerably increase the difficulty of aligning the two data sets.

The mountain data set was developed to evaluate the robustness of registration algorithms in the presence of unstructured surfaces and changes in details. It also provides a suitable test data set for multi-

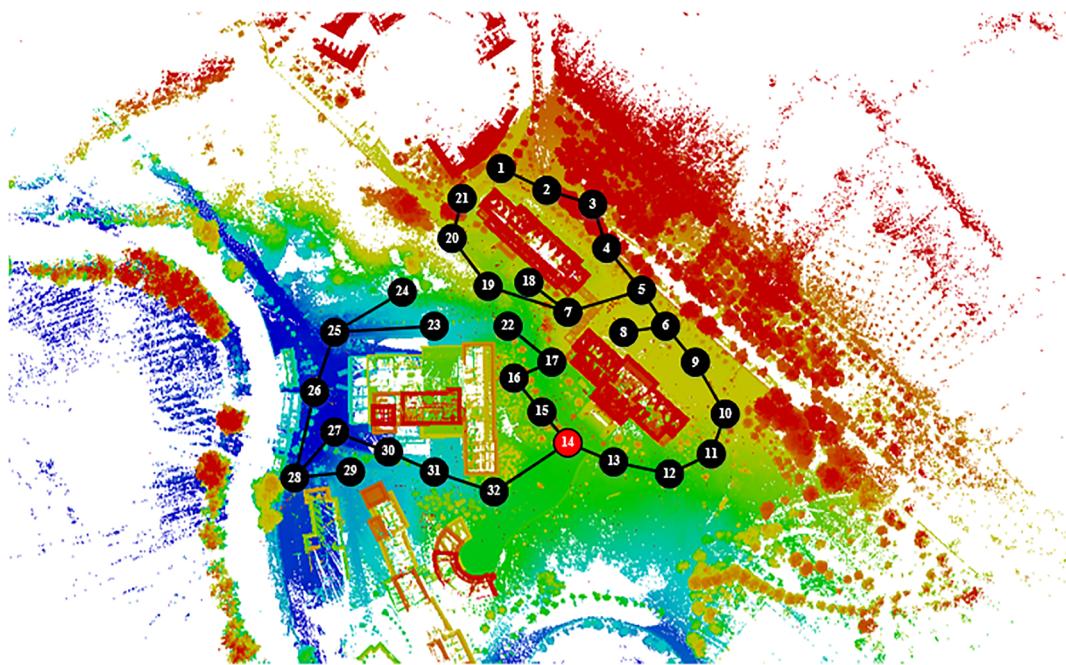


Fig. 5. Overview of the Whu-TLS park data set and registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

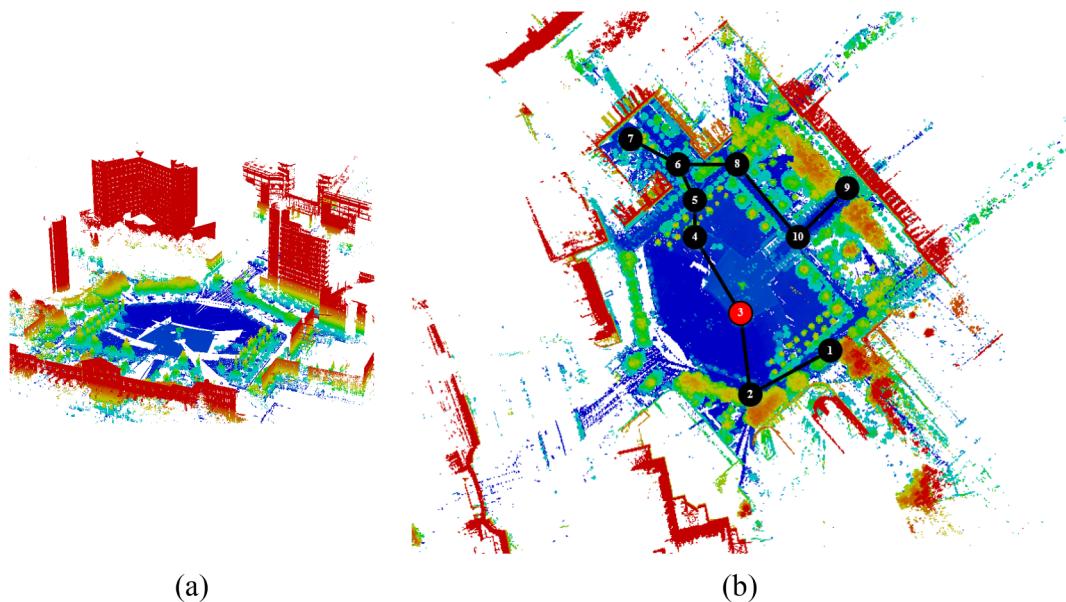


Fig. 6. Whu-TLS campus data set: (a) sampled point cloud with shading reflecting differences in height; and (b) overview of the campus data set and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

temporal registration algorithms and landslide monitoring algorithms.

[Fig. 3](#) illustrates a sampled point cloud, an overview of the whole scene, and the registration graph.

4.4. FGI-TLS forest data set

The FGI-TLS forest data set was obtained in a boreal forest sample plot measuring size $32\text{ m} \times 32\text{ m}$ located in Evo, Finland (61.19°N , 25.11°E), where Scots pine was the main tree species. The data were collected using a Leica HDS6100 TLS with a field of view of $360^\circ \times 310^\circ$ and the measurement accuracy was $\pm 2\text{ mm}$ at 25 m from the scanner. The scans were conducted in a “high density” mode, which

yielded a point spacing of 15.7 mm at 25 m in both the horizontal and vertical directions. The forest sample plot was scanned in five locations, with one scan in the center of the plot and four scans in the four quadrant directions. Thus, this data set contains five point clouds obtained from five scanning locations. The forest data set can be used to verify the validity of algorithms for complex forest environments. It also provides examples of pole-shaped objects for multi-view registration algorithms. [Fig. 4](#) illustrates a sampled point cloud, an overview of the whole scene, and the registration graph.

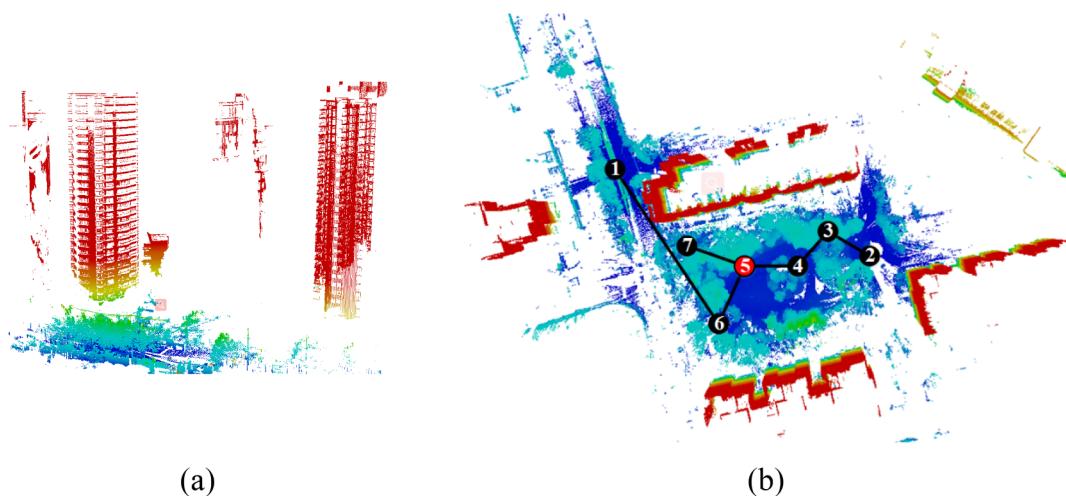


Fig. 7. Whu-TLS residence data set: (a) sampled point cloud with shading reflecting differences in height; and (b) overview of the residence data set and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

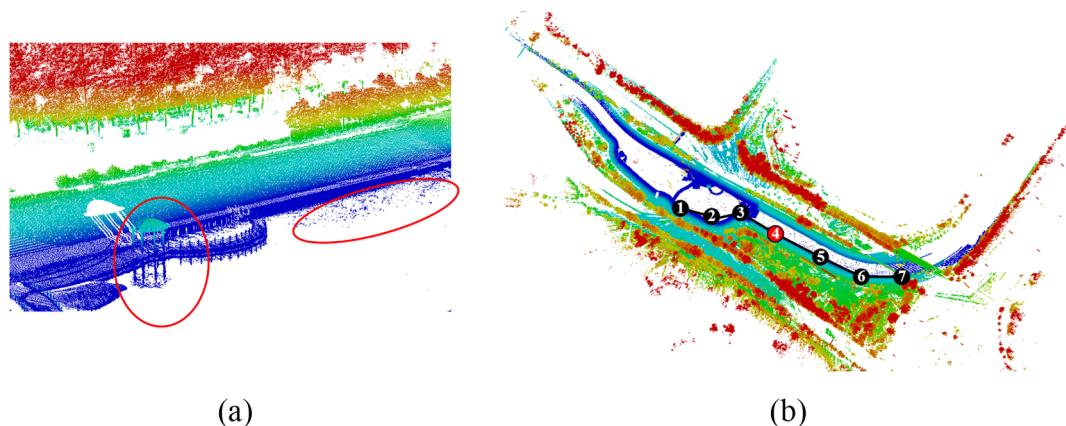


Fig. 8. Whu-TLS river bank data set: (a) sampled point cloud showing the noise caused by mirror reflection from the water (i.e., points in the red ellipse); and (b) overview of the river bank data set and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

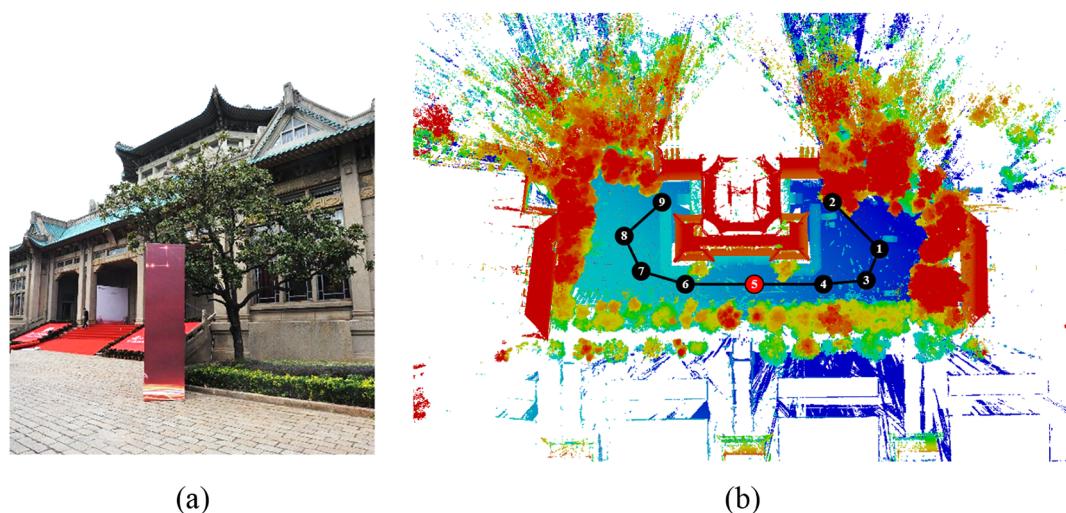


Fig. 9. Whu-TLS heritage building data set: (a) sampled image; and (b) overview of the heritage building data set and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

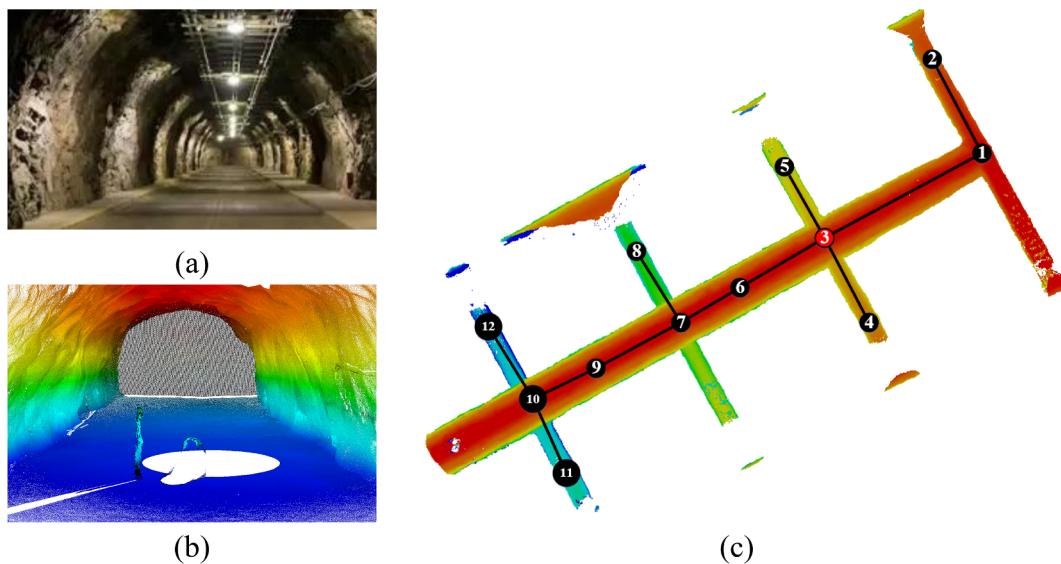


Fig. 10. Whu-TLS underground excavation data set: (a) sampled image showing the details of the underground excavation; and (b) overview of the underground excavation data set and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

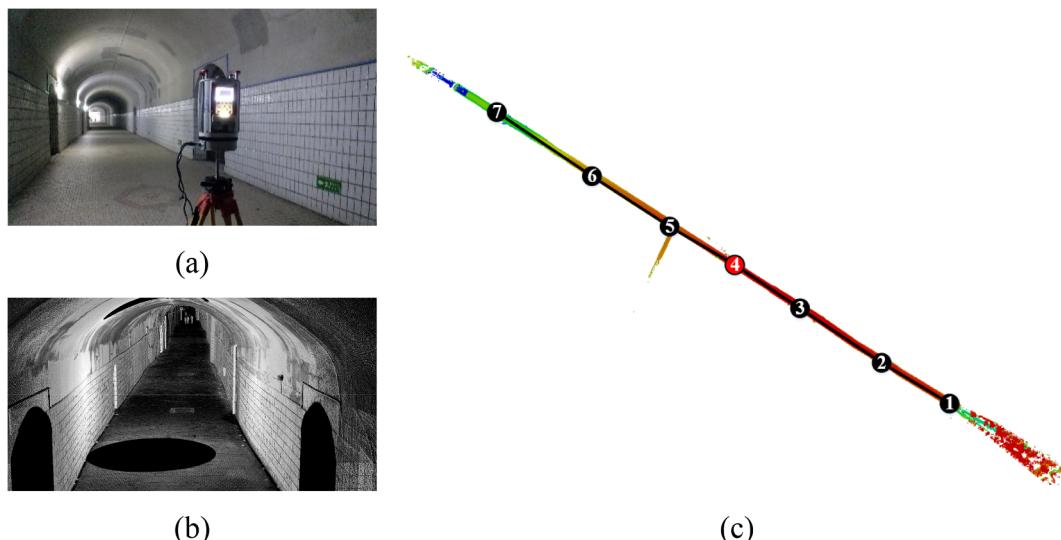


Fig. 11. Whu-TLS tunnel data set: (a) and (b) sampled photograph and point cloud showing details of the tunnel; and (c) overview of the tunnel data set and the registration graph, where the red and black circles represent the positions of anchor scans and other scans, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.5. Whu-TLS park data set

The Whu-TLS park data set is an outdoor data set containing 32 scans and 160.24 million points from a typical area with a broad view, which were captured in Longquan Park in Wuhan using the RIEGL VZ-400 laser scanner system. This data set is a representative environment containing a mixture of manmade structures (e.g., buildings, roads, and guardrail) and natural features (e.g., trees, bushes, mountain, river, and grass). The park data set can be employed to verify the validity of algorithms for use in semi-structured environments. It is also a suitable data set for large-scale multiview registration algorithms due to the availability of sufficient scans. Fig. 5 presents an overview of the whole scene and the registration graph.

4.6. Whu-TLS campus data set

The Whu-TLS campus data set is an outdoor data set containing 10

scans and 109.05 million points, which were captured in Friendship Square at Wuhan University using the Leica ScanStation P40 system. The environment contains hybrid structured elements (e.g., buildings, roads, poles, and stairs) and unstructured features (e.g., trees, sculpture, and grass). The presence of many artifacts caused by moving cars and pedestrians make this data set a suitable challenge to test the robustness of registration algorithms. Fig. 6 illustrates a sampled point cloud, an overview of the whole scene, and the registration graph.

4.7. Whu-TLS residence data set

The Whu-TLS residence data set was collected using the Leica ScanStation P40 system. We widened the distance between two adjacent scans to accelerate the data acquisition efficiency in the field, which resulted in limited overlaps between adjacent scans (e.g., the overlap between point cloud 1 and point cloud 6 was 1.0%). The residence data set contains many repetitive structures (e.g., windows) and

homogeneous architectural layouts, which add to the ambiguity during registration. Therefore, the residence data set can be used to verify the performance of registration algorithms in the presence of limited overlaps, and repetitive and symmetric elements. Fig. 7 illustrates a sampled point cloud, an overview of the whole scene, and the registration graph.

4.8. Whu-TLS riverbank data set

The Whu-TLS riverbank data set is an outdoor data set containing seven scans, 93.11 million points, and fairly high overlapping, which were captured along a river using the RIEGL VZ-400 laser scanner system. The environment contains hybrid structured elements (e.g., bridge and roads) and unstructured features (e.g., trees and grass) with a broad view. Most of the laser pulses were absorbed by the water, so the water areas are represented as data holes. Other laser pulses were reflected by the water to generate fake points (i.e., noise). The riverbank data set can be used to verify the performance of registration algorithms in the presence of data holes and noise. This data set is also suitable for river surveys and *regulation* applications. Fig. 8 illustrates a sampled point cloud showing the noise caused by mirror reflection from water (i.e., points in the red ellipse), an overview of the whole scene, and the registration graph.

4.9. Whu-TLS heritage building data set

The Whu-TLS heritage building data set (i.e., the ancient library at Wuhan University) is an outdoor data set containing nine scans, 238.16 million points, and 11 images, which were captured using the RIEGL VZ-400 laser scanner system and 5D Mark II camera. Academic activity was occurring while the scanner was recording, which resulted in many moving people. The data set provides a challenging environment containing unstructured architecture (i.e., roof pick gallery and tiles). It is also a suitable data set for reconstructing 3D models of the heritage building, cultural heritage conservation, and the fusion of point clouds and images. Fig. 9 illustrates a sampled image, an overview of the whole scene, and the registration graph.

4.10. Whu-TLS underground excavation data set

The Whu-TLS underground excavation data set containing 12 scans and 482.42 million points was collected using the RIEGL VZ-400 laser scanner system. This underground excavation environment was constructed to store chemical materials and the scans mainly contain unstructured elements (i.e., exposed rock) and structured elements (e.g., pipelines). The exposed rock has repetitive and symmetric features, thereby increasing the challenge during registration. The underground excavation data set can be used to verify the validity of algorithms for long and narrow scenes with unstructured elements. It is also a suitable data set for reconstructing 3D models of underground excavations and underground asset management applications. Fig. 10 presents a photograph of the underground excavation, an overview of the whole scene, and the registration graph.

4.11. Whu-TLS tunnel data set

The Whu-TLS tunnel data set containing seven scans and 157,018,478 points was collected using the RIEGL VZ-400 TLS system. The tunnel data set contains extremely repetitive and symmetric structures (e.g., the floor and tiles on the wall) as well as weak geometric features (e.g., roof of the tunnel), thereby adding uncertainty in the correspondence matching process. In addition, we enlarged the baseline between two scans to accelerate the data acquisition efficiency, which led to extremely low overlaps between adjacent scans (e.g., the minimum overlap is less than 6.0% between point clouds 4 and 5). Furthermore, the one-way and linear layout of the scans readily leads to

the accumulation of registration errors, which are difficult to solve using multiview registration methods. Thus, the tunnel data set is the most challenging data set for registration. Fig. 11 illustrates a photograph of the tunnel, an overview of the whole scene, and the registration graph.

5. Further research directions

In the following, we discuss possible future directions for TLS point cloud registration based on contemporary research.

5.1. Large-scale benchmark data sets

Deep learning-based methods have become the de-facto baseline for image registration tasks in the field of computer vision, but there is still no fully effective method for 3D point cloud registration due to the lack of large benchmark data sets. Large-scale benchmark data sets would allow the benchmarking of state-of-the-art algorithms in this field, as well as facilitating comparisons and providing insights into the advantages and disadvantages of different registration methods based on a common standard. Furthermore, the fusion of hyperspectral and geometric knowledge regarding benchmark data sets is expected to generate better results, particularly in situations where methods based on single geometric characteristics or single hyperspectral characteristics perform badly.

5.2. Auxiliary data-guided registration

External data or auxiliary sensors (e.g., camera and mobile phone) tend to be utilized to capture the position and orientation information in each scan (Chen et al., 2017; Ge et al., 2019). For instance, Chen et al. (2017) integrated a cut-price mobile phone with the TLS system to improve the robustness and efficiency of multiview registration. In particular, the locations of the scanner and the distances between the surrounding scanners were approximately determined using the GPS receiver in the smartphone and employed as constraints for multiview registration. Ge et al. (2019) introduced an image-driven registration method for global TLS point cloud alignment and their results indicated that the incorporation of panoramic images can contribute greatly to both coarse and fine registration. Advanced TLS systems (e.g., Leica RTC360) can automatically record the motions from station to station in order to pre-register scans in the field using the SLAM technique, and the edge computing technique is utilized for automatically registering scans in real-time.¹²

5.3. Deep learning-based registration

Due to the development of computer vision techniques, deep learning-based methods have achieved great success and provided a new approach for solving TLS point cloud registration problems. In contrast to handcrafted feature-based methods, deep learning-based methods can jointly learn the keypoint detection, feature representation, and correspondence matching processes in an end-to-end manner, thereby enhancing the capacity for registering challenging scenarios containing extremely repetitive and symmetric elements, poor geometric features, minimal overlaps, and scenario changes. The existing deep learning-based registration methods perform well at small-scale indoor point cloud registration (Deng et al., 2018a, 2018b; Yew and Lee, 2018; Deng et al., 2019; Lu et al., 2019; Wang and Solomon, 2019) but to the best of our knowledge, the deep learning-based registration methods for large-scale outdoor TLS point clouds are still in the early stage of development because of their limited capacity to deal with

¹² <https://leica-geosystems.com/products/laser-scanners/scanners/leica-rtc360>

large volumes and irregular point clouds.

5.4. Multi-temporal point cloud registration

The registration of multi-temporal point clouds should also be investigated because accurate registration and alignment with the multi-temporal data set is essential for applications such as change detection. However, the dynamics of the scene, especially changed details, will considerably increase the difficulty of aligning two data sets. To some extent, these dynamics are more problematic than low overlap because the differences are highly likely to cause correspondence mismatches. One possible solution to this problem involves eliminating the dynamic objects (e.g., pedestrians and moving vehicles) and details (e.g., small objects and ornaments), and simply aligning the two data sets with fundamental structures (e.g., layouts of buildings or landmarks), which are normally permanent and stable during the entire observation period. In practical applications, separating these fundamental structures can be achieved more easily in the phase domain than the Euclidean space because the fundamental structures generally correspond to low frequency components in the phase domain and they are simple to extract (Huang et al., 2019).

6. Conclusion

The registration of unsorted TLS point clouds is essential for many tasks in computer vision, photogrammetry, and remote sensing domains. In this study, we comprehensively reviewed TLS point cloud registration in terms of pairwise coarse registration, pairwise fine registration, and multiview registration, and identified the strengths and weaknesses of the methods employed. In addition, we proposed a new benchmark data set containing a wide range of scenes (i.e., subway station, high-speed railway platform, mountain, forest, park, campus, residence, river bank, heritage building, underground excavation, and tunnel environments) with significant differences in terms of the organization of the environment and geometric shapes in order to facilitate the development of cutting-edge TLS point cloud registration methods. Finally, we summarized some practical problems that affect registration as well as providing our own observations, insights, and suggestions for future research. We hope that this survey provides a useful comparison of the strengths and weaknesses of the existing algorithms, and that the proposed benchmark provides an important data set to support the development of deep learning-based TLS point cloud registration methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was jointly supported by the National Natural Science Foundation of China Projects (Nos. 41901403, 41725005, and 41531177) and the China Postdoctoral Science Found (No. 2018M642913). The authors would like to acknowledge the Finnish Geospatial Research Institute for providing the forest data set.

References

- Aiger, D., Mitra, N., Cohen-Or, D., 2008. Four-points congruent sets for robust surface registration. *ACM Trans. Graphics* 27 (3), 1–10.
- Andres, B., Kappes, J. H., Beier, T., Köthe, U., & Hamprecht, F. A., 2012. The lazy flipper: efficient depth-limited exhaustive search in discrete graphical models. In: European Conference on Computer Vision. Springer, Berlin, October. 2012, 154–166.
- Aoki, Y., Goforth, H., Srivatsan, R. A., Lucey, S., 2019. PointNetLK: Robust & efficient point cloud registration using PointNet. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7163–7172.
- Arrigoni, F., Rossi, B., Fusello, A., 2016. Global registration of 3D point sets via LRS decomposition. In: European Conference on Computer Vision. Springer, Cham, October. 2016, 489–504.
- Bae, K.H., Lichten, D.D., 2008. A method for automated registration of unorganized point clouds. *ISPRS J. Photogramm. Remote Sens.* 63 (1), 36–54.
- Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 239–256.
- 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), vol. 3, 2743–2748.
- Boerner, R., Xu, Y., Baran, R., Steinbacher, F., Hoegner, L., Stilla, U., 2019. Registration of multi-sensor bathymetric point clouds in rural areas using point-to-grid distances. *ISPRS Int. Geo-Inf.* 8 (4), 178.
- Bouaziz, S., Tagliasacchi, A., Pauly, M., 2013. Ppfnet: Global context aware local features for robust 3d point matching. In: Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing. Eurographics Association. 113–123.
- Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 1222–1239.
- Bueno, M., Bosché, F., González-Jorge, H., Martínez-Sánchez, J., Arias, P., 2018. 4-Plane congruent sets for automatic registration of as-is 3D point clouds with 3D BIM models. *Autom. Constr.* 89, 120–134.
- Cai, Z., Chin, T.J., Bustos, A.P., Schindler, K., 2019. Practical optimal registration of terrestrial LiDAR scan pairs. *ISPRS J. Photogramm. Remote Sens.* 147, 118–131.
- Campbell, R.J., Flynn, P.J., 2001. A survey of free-form object representation and recognition techniques. *Comput. Vis. Image Underst.* 81 (2), 166–210.
- Chen, M., Wang, S., Wang, M., Wan, Y., He, P., 2017. Entropy-based registration of point clouds using terrestrial laser scanning and smartphone GPS. *Sensors* 17 (1), 197.
- Chen, Y., Medioni, G., Object modeling by registration of multiple range images. In: Proceedings of the IEEE International Conference on Robotics and Automation, Sacramento, CA, USA, April. 1991, 2724–2729.
- Cheng, L., Chen, S., Liu, X., Xu, H., Wu, Y., Li, M., Chen, Y., 2018. Registration of laser scanning point clouds: a review. *Sensors* 18 (5), 1641.
- Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P., 2002. The trimmed iterative closest point algorithm. In: IEEE Object Recognition Supported by User Interaction for Service Robots, 3, 545–548.
- Das, A., Waslander, S.L., 2012. Scan registration with multi-scale k-means normal distributions transform. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2705–2710.
- Das, A., Waslander, S.L., 2014. Scan registration using segmented region growing NDT. *Int. J. Robot. Res.* 33 (13), 1645–1663.
- Deng, H., Birdal, T., & Ilic, S., 2018b. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In: European Conference on Computer Vision. Springer, Munich, September. 2018, 602–618.
- Deng, H., Birdal, T., Ilic, S., 2018a. 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.
- Deng, H., Birdal, T., Ilic, S., 2019. 3D Local features for direct pairwise registration. arXiv preprint arXiv:1904.04281.
- Dold, C., Brenner, C., 2006. Registration of terrestrial laser scanning data using planar patches and image data. *Int. Arch. Photogram. Remote Sens. Spatial Info. Sci.* 36 (5), 78–83.
- Dong, J., Peng, Y., Ying, S., Hu, Z., 2014. LieTriCP: An improvement of trimmed iterative closest point algorithm. *Neurocomputing* 140, 67–76.
- Dong, Z., Yang, B., Hu, P., Scherer, S., 2018a. An efficient global energy optimization approach for robust 3D plane segmentation of point clouds. *ISPRS J. Photogramm. Remote Sens.* 137, 112–133.
- Dong, Z., Yang, B., Liang, F., Huang, R., Scherer, S., 2018b. Hierarchical registration of unsorted TLS point clouds based on binary shape context descriptor. *ISPRS J. Photogramm. Remote Sens.* 144, 61–79.
- Dong, Z., Yang, B., Liu, Y., Liang, F., Li, B., Zang, Y., 2017. A novel binary shape context for 3d local surface description. *ISPRS J. Photogramm. Remote Sens.* 130, 431–452.
- Evangelidis, G.D., Horaud, R., 2017. Joint alignment of multiple point sets with batch and incremental expectation-maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (6), 1397–1410.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24 (6), 381–395.
- Fitzgibbon, A.W., 2003. Robust registration of 2D and 3D point sets. *Image Vis. Comput.* 21 (13–14), 1145–1153.
- Freitas, S., Catita, C., Redweik, P., Brito, M.C., 2015. Modelling solar potential in the urban environment: state-of-the-art review. *Renew. Sust. Energ. Rev.* 41, 915–931.
- Ge, X., Wunderlich, T., 2016. Surface-based matching of 3D point clouds with variable coordinates in source and target system. *ISPRS J. Photogramm. Remote Sens.* 111, 1–12.
- Ge, X., 2017a. Automatic markerless registration of point clouds with semantic-keypoint-based 4-points congruent sets. *ISPRS J. Photogramm. Remote Sens.* 130, 344–357.
- Ge, X., 2017b. Non-rigid registration of 3D point clouds under isometric deformation. *ISPRS J. Photogramm. Remote Sens.* 121, 192–202.
- Ge, X., Hu, H., Wu, B., 2019. Image-guided registration of unsorted terrestrial laser scanning point clouds for urban scenes. *IEEE Trans. Geosci. Remote Sens.* 57 (11), 9264–9276.
- Ge, X., Hu, H., 2020. Object-based incremental registration of terrestrial point clouds in an urban environment. *ISPRS J. Photogramm. Remote Sens.* 161, 218–232.
- Gehrung, J., Hebel, M., Arens, M., Stilla, U., 2016. A framework for voxel-based global

- scale modeling of urban environments. *Int. Arch. Photogram. Remote Sens. Spatial Info Sci* 42.
- Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? The kitti vision benchmark suite. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 3354–3361.
- Glent Buch, A., Yang, Y., Kruger, N., Gordon Petersen, H., 2014. In search of inliers: 3d correspondence by local and global voting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2067–2074.
- Golyanik, V., Taetz, B., Reis, G., Stricker, D., 2016. Extended coherent point drift algorithm with correspondence priors and optimal subsampling. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), March. 2016, 1–9.
- Govindu, V.M., Pooja, A., 2013. On averaging multiview relations for 3d scan registration. *IEEE Trans. Image Process.* 23 (3), 1289–1302.
- Gressin, A., Mallet, C., Demantké, J., David, N., 2013. Towards 3D lidar point cloud registration improvement using optimal neighborhood knowledge. *ISPRS J. Photogramm. Remote Sens.* 79, 240–251.
- Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J., Kwok, N.M., 2014. An accurate and robust range image registration algorithm for 3d object modeling. *IEEE Transactions on Multimedia* 1377–1390.
- Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J., Kwok, N.M., 2016. A comprehensive performance evaluation of 3D local feature descriptors. *Int. J. Comput. Vision* 116 (1), 66–89.
- Guo, Y., Sohel, F., Bennamoun, M., Lu, M., Wan, J., 2013. Rotational projection statistics for 3D local surface description and object recognition. *Int. J. Comput. Vision* 105 (1), 63–86.
- Guo, Y., Sohel, F., Bennamoun, M., Wan, J., Lu, M., 2015. A novel local surface feature for 3D object recognition under clutter and occlusion. *Inf. Sci.* 293 (2), 196–213.
- Habib, A., Detchev, I., Bang, K., 2010. A comparative analysis of two approaches for multiple-surface registration of irregular point clouds. *Int. Arch. Photogram. Remote Sens. Spatial Info. Sci.* 38 (1), 61–66.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Hebel, M., Stilla, U., 2012. Simultaneous calibration of ALS systems and alignment of multiview LiDAR scans of urban areas. *IEEE Trans. Geosci. Remote Sens.* 50 (6), 2364–2379.
- Huang, J., Kwok, T.-H., Zhou, C., 2017. V4PCS: Volumetric 4PCS algorithm for global registration. *J. Mech. Des.* 139 (11).
- 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. *Int. Arch. Photogram. Remote Sens. Spatial Info. Sci.* 42, 2/W13.
- Huber, D.F., Hebert, M., 2003. Fully automatic registration of multiple 3D data sets. *Image Vis. Comput.* 21 (7), 637–650.
- Huttenlocher, D.P., 1991. Fast affine point matching: an output-sensitive method. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 263–268.
- Jegou, H., Perronnin, F., Douze, M., Sánchez, J., Perez, P., Schmid, C., 2012. Aggregating local image descriptors into compact codes. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (9), 1704–1716.
- Jian, B., Vemuri, B.C., 2010. Robust point set registration using Gaussian mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (8), 1633–1645.
- Johnson, A.E., Sing, B.K., 1999. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transact. Pattern Anal. Mach. Intell.*
- Jung, J., Hong, S., Jeong, S., Kim, S., Cho, H., Hong, S., Heo, J., 2014. Productive modeling for development of as-built BIM of existing indoor structures. *Autom. Constr.* 42, 68–77.
- Kelbe, D., Van, A.J., Romanczyk, P., Van, L.M., Cawse-Nicholson, K., 2016. Marker-free registration of forest terrestrial laser scanner data pairs with embedded confidence metrics. *IEEE Trans. Geosci. Remote Sens.* 54 (7), 4314–4330.
- Khoshelham, K., Nardinocchi, C., Frontoni, E., Mancini, A., Zingaretti, P., 2010. Performance evaluation of automated approaches to building detection in multi-source aerial data. *ISPRS J. Photogramm. Remote Sens.* 65 (1), 123–133.
- Kruskal, J.B., 1956. On the shortest spanning subtree of a graph and the traveling salesmen problem. *Proc. Am. Math. Soc.* 7 (1), 48–50.
- Lawin, F.J., Daneljan, M., Khan, F.S., Forssén, P.E., Felsberg, M., 2018. Density adaptive point set registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 3829–3837.
- Liang, X., Litkey, P., Hyppä, J., Kaartinen, H., Vastaranta, M., Holopainen, M., 2012. Image-guided registration of unorganized terrestrial laser scanning point clouds for urban scenes. *IEEE Trans. Geosci. Remote Sens.* 50 (2), 661–670.
- Lu, M., Zhao, J., Guo, Y., Ou, J., Li, J., 2014. A 3D point cloud registration algorithm based on fast coherent point drift. In: 2014 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), October. 2014, 1–6.
- Lu, W., Wan, G., Zhou, Y., Fu, X., Yuan, P., Song, S., 2019. Deep ICP: An end-to-end deep neural network for 3D point cloud registration. arXiv preprint arXiv:1905.04153.
- Magnusson, M., Duckett, T., 2005. A comparison of 3d registration algorithms for autonomous underground mining vehicles. In: Proceedings of the European Conference on Mobile Robotics (ECMR 2005). 86–91.
- Magnusson, M., 2009. The three-dimensional normal-distributions transform: an efficient representation for registration, surface analysis, and loop detection (Doctoral dissertation, Örebro universitet).
- Magnusson, M., Lilienthal, A., Duckett, T., 2007. Scan registration for autonomous mining vehicles using 3D-NDT. *J. Field Rob.* 24 (10), 803–827.
- Mellado, N., Aiger, D., Mitra, N., 2014. Super 4PCS fast global pointcloud registration via smart indexing. *Comput. Graph. Forum* 33 (5), 205–215.
- Mian, A., Bennamoun, M., Owens, R., 2006. Three dimensional model-based object recognition and segmentation in cluttered scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (10), 1584–1601.
- Min, Z., Meng, M.Q.H., 2019. Robust generalized point set registration using inhomogeneous hybrid mixture models via expectation maximization. 2019 International Conference on Robotics and Automation (ICRA) 8733–8739.
- Mohamad, M., Ahmed, M., Rappaport, D., Greenspan, M., 2015. Super generalized 4PCS for 3D registration. In: International Conference on 3D Vision, pp. 598–606.
- Mohamad, M., Rappaport, D., Greenspan, M., 2014. Generalized 4-points congruent sets for 3d registration. In: 2nd International Conference on 3D Vision, pp. 83–90.
- Montuori, A., Luzi, G., Stramondo, S., Casula, G., Bignami, C., Bonali, E., Bianchi, M.G., Crosetto, M., 2014. Combined use of ground-based systems for Cultural Heritage conservation monitoring. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS). 4086–4089.
- Morgan, M., Habib, A., 2002. Interpolation of lidar data and automatic building extraction. In: Proceedings of ACSM-ASPRS Annual Conference, April, 2002, 432–441.
- Myronenko, A., Song, X., 2010. Point set registration: coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (12), 2262–2275.
- Myronenko, A., Song, X., Carreira-Perpinán, M.A., 2007. Non-rigid point set registration: Coherent point drift. In: Advances in Neural Information Processing Systems, 1009–1016.
- Nüchter, A., 2008. 3D Robotic Mapping: the Simultaneous Localization and Mapping Problem with Six Degrees of Freedom. Springer.
- Oesau, S., Lafarge, F., Alliez, P., 2014. Indoor scene reconstruction using feature sensitive primitive extraction and graph-cut. *ISPRS J. Photogramm. Remote Sens.* 90, 68–82.
- Papadopoulos, T., Lourakis, M. I., 2000. Estimating the Jacobian of the singular value decomposition: Theory and applications. In: European Conference on Computer Vision. Springer, Berlin, June. 2000, 554–570.
- Pavlov, A.L., Ovchinnikov, G.W., Derbyshev, D.Y., Tsetserukou, D., Oseledets, I.V., 2018. AA-ICP: Iterative closest point with Anderson acceleration. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), May. 2018, 1–6.
- Pieraccini, M., Noferini, L., Mecatti, D., Atzeni, C., Teza, G., Galgaro, A., Zaltron, N., 2006. Integration of radar interferometry and laser scanning for remote monitoring of an urban site built on a sliding slope. *IEEE Trans. Geosci. Remote Sens.* 44 (9), 2335–2342.
- Polewski, P., Yao, W., Heurich, M., Krzystek, P., Stilla, U., 2018. Learning a constrained conditional random field for enhanced segmentation of fallen trees in ALS point clouds. *ISPRS J. Photogramm. Remote Sens.* 140, 33–44.
- Pomerleau, F., Colas, F., Siegwart, R., 2015. A review of point cloud registration algorithms for mobile robotics. *Found. Trends® Robot.* 4(1), 1–104.
- Pomerleau, F., Krusi, P., Colas, F., Furgale, P., & Siegwart, R., 2014. Long-term 3D map maintenance in dynamic environments. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), May. 2014, 3712–3719.
- Prokop, A., Panholzer, H., 2009. Assessing the capability of terrestrial laser scanning for monitoring slow moving landslides. *Nat. Hazards Earth Syst. Sci.* 9 (6), 1921–1928.
- Pujol-Miró, A., Casas, J.R., Ruiz-Hidalgo, J., 2019. Correspondence matching in unorganized 3D point clouds using Convolutional Neural Networks. *Image Vis. Comput.* 83, 563–570.
- Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660.
- Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems. 5099–5108.
- Qi, D., May, S., Nüchter, A., 2009. GPU-accelerated nearest neighbor search for 3D registration. In: International Conference on Computer Vision Systems. Springer, Berlin, October. 2009, 194–203.
- Rabbani, T., Dijkman, S., van den Heuvel, F., Vosselman, G., 2007. An integrated approach for modelling and global registration of point clouds. *ISPRS J. Photogramm. Remote Sens.* 61 (6), 355–370.
- Redweik, P., Catita, C., Brito, M., 2013. Solar energy potential on roofs and facades in an urban landscape. *Sol. Energy* 97, 332–341.
- Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. In: 2009 IEEE International Conference on Robotics and Automation (ICRA), May. 2009, 3212–3217.
- Salvi, J., Matabosch, C., Fofi, D., Forest, J., 2007. A review of recent range image registration methods with accuracy evaluation. *Image Vis. Comput.* 25 (5), 578–596.
- Segal, A., Haehnel, D., Thrun, S., 2009. Generalized-icp. *Robot.: Sci. Syst.* 5, 168–176.
- Sharp, G.C., Lee, S.W., Wehe, D.K., 2002. ICP registration using invariant features. *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (1), 90–102.
- Shih, S.W., Chuang, Y.T., Yu, T.Y., 2008. An efficient and accurate method for the relaxation of multiview registration error. *IEEE Trans. Image Process.* 17 (6), 968–981.
- Simon, D.A., Hebert, M., Kanade, T., 1995. Techniques for fast and accurate intrasurgical registration. *Comput. Aided Surg.* 1 (1), 17–29.
- Stamos, I., & Leordeanu, M., 2003. Automated feature-based range registration of urban scenes of large scale. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1063–6919.
- Stoyanov, T., Magnusson, M., Andreasson, H., Lilienthal, A.J., 2012. Fast and accurate scan registration through minimization of the distance between compact 3D NDT representations. *Int. J. Robot. Res.* 31 (12), 1377–1393.
- Takeuchi, E., Tsubouchi, T., 2006. A 3-D scan matching using improved 3-D normal distributions transform for mobile robotic mapping. In: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3068–3073.
- Tam, G.K., Cheng, Z.Q., Lai, Y.K., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X., Rosin, P.L., 2012. Registration of 3D point clouds and meshes: a survey from rigid to nonrigid. *IEEE Trans. Vis. Comput. Graph.* 19 (7), 1199–1217.
- Tang, Y., Feng, J., 2015. Hierarchical multiview rigid registration. *Comput. Graph. Forum*

- 34 (5), 77–87.
- Tazir, M. L., Gokhool, T., Checchin, P., Malaterre, L., & Trassoudaine, L., 2018. Cluster ICP: Towards Sparse to Dense Registration. In: International Conference on Intelligent Autonomous Systems. Springer, Cham, June. 2018, 730–747.
- Theiler, P., Wegner, J., Schindler, K., 2014a. Fast registration of laser scans with 4-point congruent sets: what works and what doesn't. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* 2 (3), 149–156.
- Theiler, P., Wegner, J., Schindler, K., 2014b. Keypoint-based 4-points congruent sets: Automated marker-less registration of laser scans. *ISPRS J. Photogramm. Remote Sens.* 96, 149–163.
- Tombari, F., Salti S., Distefano L., 2010. Unique signatures of histograms for local surface description. In: European Conference on Computer Vision. Springer, Berlin, September. 2010, 356–369.
- Theiler, P.W., Wegner, J., Schindler, K., 2015. Globally consistent registration of terrestrial laser scans via graph optimization. *ISPRS J. Photogramm. Remote Sens.* 126–138.
- Tombari, F., Salti, S., Di Stefano, L., 2013. Performance evaluation of 3D keypoint detectors. *Int. J. Comput. Vision* 102 (1–3), 198–220.
- Tsin, Y., Takeo, K., 2004. A correlation-based approach to robust point set registration. In: European Conference on Computer Vision. Springer, Springer, Berlin, 2004, 558–569.
- Tuttas, S., Braun, A., Borrman, A., Stilla, U., 2017. Acquisition and consecutive registration of photogrammetric point clouds for construction progress monitoring using a 4D BIM. *PFG-J. Photogramm. Remote Sens. Geoinf. Sci.* 85 (1), 3–15.
- Uhlenbrock, R., Kim, K., Hoffmann, H., Dolne, J., 2017. Rapid 3D registration using local subtree caching in iterative closest point (ICP) algorithm. In: Unconventional and Indirect Imaging, Image Reconstruction, and Wavefront Sensing 2017. International Society for Optics and Photonics. 2017, September. 10410: 10410J.
- Vosselman, G., Mass, H.-G., 2010. Airborne and Terrestrial Laser Scanning. Whittles Publishing, Scotland, UK.
- Wang, P., Wang, P., Qu, Z., Gao, Y., Shen, Z., 2011. A refined coherent point drift (CPD) algorithm for point set registration. *Sci. China Inf. Sci.* 54 (12), 2639–2646.
- Wang, Y., Solomon, J.M., 2019. Deep Closest Point: Learning Representations for Point Cloud Registration. arXiv preprint arXiv:1905.03304.
- Weber, T., Hänsch, R., Hellwich, O., 2015. Automatic registration of unordered point clouds acquired by Kinect sensors using an overlap heuristic. *ISPRS J. Photogramm. Remote Sens.* 102, 96–109.
- Weinmann, M., 2016. Reconstruction and Analysis of 3D Scenes: From Irregularly Distributed 3D Points to Object Classes. Springer.
- Xiao, J., Zhang, J., Adler, B., Zhang, H., Zhang, J., 2013. Three-dimensional point cloud plane segmentation in both structured and unstructured environments. *Rob. Auton. Syst.* 61 (12), 1641–1652.
- 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 Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* 4, 185.
- 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 J. Photogramm. Remote Sens.* 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 J. Photogramm. Remote Sens.* 113, 43–58.
- Yang, B., Zang, Y., 2014. Automated registration of dense terrestrial laser-scanning point clouds using curves. *ISPRS J. Photogramm. Remote Sens.* 95, 109–121.
- Yang, J., Li, H., Jia, Y., 2013. Go-ICP: solving 3D registration efficiently and globally optimally. In: Proceedings of the IEEE Conference on Computer Vision, pp. 1457–1464.
- Yang, Y., Feng, C., Shen, Y., Tian, D., 2018. Foldingnet: Point cloud auto-encoder via deep grid deformation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 206–215.
- Yew, Z. J., Lee, G. H., 2018. 3DFeat-Net: Weakly supervised local 3D features for point cloud registration. In: European Conference on Computer Vision. Springer, Cham, September. 2018, 630–646.
- Yu, F., Xiao, J., Funkhouser, T., 2015. Semantic alignment of LiDAR data at city scale. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1722–1731.
- Zai, D., Li, J., Guo, Y., Cheng, M., Huang, P., Cao, X., Wang, C., 2017. Pairwise registration of TLS point clouds using covariance descriptors and a non-cooperative game. *ISPRS J. Photogramm. Remote Sens.* 134, 15–29.
- Zang, Y., Lindenbergh, R., 2019. An improved coherent point drift method for TLS point cloud registration of complex scenes. *Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci.* 42(2/W13).
- Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., & Funkhouser, T., 2017. 3dmatch: Learning local geometric descriptors from rgbd reconstructions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1802–1811.
- Zhang, Z., Sun, L., Zhong, R., Chen, D., Xu, Z., Wang, C., Qin, C., Sun, H., Li, R., 2019. 3-D deep feature construction for mobile laser scanning point cloud registration. *IEEE Geosci. Remote S.* 16 (12), 1904–1908.
- Zhou, L., Zhu, S., Luo, Z., Shen, T., Zhang, R., Zhen, M., Fang, T., Quan, L., 2018. Learning and matching multi-view descriptors for registration of point clouds. In: European Conference on Computer Vision. Springer, September. 2018, Munich, 505–522.
- Zhu, J., Jiang, Z., Evangelidis, G.D., Zhang, C., Pang, S., Li, Z., 2019. Efficient registration of multi-view point sets by K-means clustering. *Inf. Sci.* 488, 205–218.
- Zhu, J., Zhu, L., Jiang, Z., Bai, X., Li, Z., Wang, L., 2017. Local to global registration of multi-view range scans using spanning tree. *Comput. Electr. Eng.* 58 (1), 477–488.
- Zhu, J., Zhu, L., Li, Z., Li, C., Cui, J., 2016. Automatic multi-view registration of un-ordered range scans without feature extraction. *Neurocomputing* 171, 1444–1453.