ISPRS Journal of Photogrammetry and Remote Sensing 130 (2017) 162–186

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Automatic co-registration of 3D multi-sensor point clouds

Ravi Ancil Persad *, Costas Armenakis

Geomatics Engineering, GeoICT Lab, Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, 4700 Keele St., Toronto, Ontario M3J 1P3, Canada

ARTICLE INFO

Article history:
Received 27 January 2017
Received in revised form 19 May 2017
Accepted 19 May 2017

Keywords:
Keypoints
Wavelets
Descriptors
Matching
Gabor filter
Rapid transform
UAV
Laser-scanning
DSM
Co-registration
Automation

ABSTRACT

We propose an approach for the automatic coarse alignment of 3D point clouds which have been acquired from various platforms. The method is based on 2D keypoint matching performed on height map images of the point clouds. Initially, a multi-scale wavelet keypoint detector is applied, followed by adaptive non-maxima suppression. A scale, rotation and translation-invariant descriptor is then computed for all keypoints. The descriptor is built using the log-polar mapping of Gabor filter derivatives in combination with the so-called Rapid Transform. In the final step, source and target height map keypoint correspondences are determined using a bi-directional nearest neighbour similarity check, together with a threshold-free modified-RANSAC. Experiments with urban and non-urban scenes are presented and results show scale errors ranging from 0.01 to 0.03, 3D rotation errors in the order of 0.2° to 0.3° and 3D translation errors from 0.09 m to 1.1 m.

1. Introduction

Automatic 3D point cloud registration is a major research topic in photogrammetry, computer vision and computer graphics. In many instances, there is the need to align point cloud data collected at different times from different platforms including laser scanners (e.g., aerial, terrestrial and mobile scanners), satellite systems and unmanned aerial vehicles (UAV). Fusion of multi-sensory data has numerous applications in 3D building modelling and reconstruction, change detection and map-revision in urban and non-urban environments, crime scene reconstruction, and mapping of open-pit mines.

There are two main phases for pairwise 3D point cloud registration: (i) the *initial, coarse* alignment, and (ii) the *refined* alignment. Both require the computation of a mathematical mapping between two point cloud datasets. This mapping is used to transform the 'source' point cloud to the 'target' point cloud. For over two decades, the refined alignment problem has received considerable attention since the development of the influential 'Iterative Closest Point' (ICP) algorithm (Besl and McKay, 1992; Chen and Medioni, 1992). Rusinkiewicz and Levoy (2001) provide an overview of many ICP variants. Bouaziz et al. (2013) developed the so-called 'Sparse ICP' which is less sensitive to outliers than the classical ICP. In the photogrammetric community, Gruen and Akca (2005) proposed an alternative to the ICP referred to as 'Least Squares 3D Surface Matching'. Instead of using closest point for correspondences as done in ICP, Bae and Lichti (2008) developed the 'Geometric Primitive ICP' method which instead used normal vector information together with change in surface curvature for point cloud matching.

As the name indicates, the refined alignment step is applied to fine-tune the matching of a point cloud pair already assumed to be coarsely co-registered. Generally, the performance of refined alignment methods depends on the quality of matching achieved in the previous coarse alignment step. Inaccurate initial registration can lead to wrong, local minima solutions during the fine-tuning process. Motivated by this, we concentrate on addressing the initial, coarse point cloud co-registration problem.

2. Related works on automated coarse point cloud alignment

By coarse alignment we assume that there is no prior knowledge of the 3D conformal transformation parameters (i.e., single global scale factor, 3D rotation angles and 3D translations). However, in some of the reviewed literature, the scale factor is assumed to be known and only the six rigid parameters are

* Corresponding author.
E-mail addresses: ravi071@yorku.ca (R.A. Persad), armenc@yorku.ca (C. Armenakis).

http://dx.doi.org/10.1016/j.isprsjprs.2017.05.014
0924-2716/

considered as the unknowns to be computed. Instances of such cases will be identified as we proceed. There are various approaches one can apply to achieve initial source to target point cloud co-registration. We classify these into three categories: (i) 3D descriptor-based methods, (ii) 3D non-descriptor-based- methods and (iii) 2D image-based methods.

2.1. 3D descriptor-based methods

Descriptor-based methods are typically applied in 3D feature matching frameworks. They usually rely on the extraction of salient key-features (e.g., 3D keypoints) on the point cloud surface. For these keypoints, descriptors are formed by utilizing various types of local neighbourhood shape attributes of the point cloud. Similar descriptors on source and target point clouds can then be matched to find corresponding keypoints. Various 3D point cloud descriptors have been developed over the years. Some of these include the Spin Images (Johnson and Hebert, 1999), Fast Point Feature Histograms (Rusu et al., 2009) and Signature of Histograms of Orientations (Tombari et al., 2010). These descriptors require a local point cloud neighbourhood to be defined around the keypoint. A user-specified distance is applied to define local neighbourhoods when the source and target point clouds have the same scale. In situations where there is a scale difference these descriptors are not scale-invariant and will fail during the feature matching process.

Typically, scale-invariance is achieved during the keypoint detection phase, where a local scale value is computed for every keypoint. Then this local scale is used to define the local region used for descriptor generation. This concept is popularly applied for 2D keypoint descriptors. For example, the Scale Invariant Feature Transform (SIFT) detector (Lowe, 2004) uses a 'Difference-of-Gaussian' (DOG) framework for estimating the local scale, whereas the Harris-Laplacian interest point operator (Mikolajczyk and Schmid, 2004) uses Lindeberg's automatic scale selection approach (Lindeberg, 1998). There are 3D extensions of the SIFT (Flitton et al., 2010) and Speeded Up Robust Features (SURF) (Knopp et al., 2010). However, these are volume-based methods which utilize 3D voxel representations instead of direct point cloud data. In recent work, Mellado et al. (2016) developed an approach for scale-invariant co-registration of multi-sensor point clouds based on a descriptor known as 'Growing Least Squares' (GLS). The GLS descriptor is built in a logarithmic scale space, facilitating the provision of local and global scale-invariant point cloud attributes.

2.2. 3D non-descriptor-based methods

There are also descriptor-free approaches which address the coarse 3D point cloud alignment problem. A common approach for global co-registration is the utilization of Principal Component Analysis (PCA). PCA is used to approximate the rotation required to align the coordinate systems of the source and target point clouds. The translation can be estimated by the difference in centroids of the source and target data. However, when there is partial overlap and/or shape deformation between the source and target surfaces this approach does not provide the correct transformation parameters.

Other non-descriptor based methods utilize various geometric constraints and relationships amongst points, lines or planes. In terms of the plane-based methods, von Hansen (2006) presented a framework for terrestrial laser scanning (TLS) co-registration. Firstly, planes are extracted from point cloud data and this is followed by an exhaustive search for corresponding planes. The method does not cater for scale differences between the point clouds. Brenner et al. (2008) derived two methods for the coarse registration of TLS data: a plane-based scoring approach and another which uses the normal distributions transform (NDT) (Biber, 2003). In the first method, plane triplet correspondences are scored using the similarity of their normal vector directions, in combination with distances to the plane origin. The second method sliced the 3D scans into 2D layers, and then used the 2D NDT algorithm for co-registration. Only the 3D roto-translational parameters were accounted for in their work.

Stamos and Leordeanu (2003) used both linear and planar features to align laser scans of buildings. Properties such as length of the lines, in addition to plane sizes were used to discard possible erroneous matches, thus reducing the combinatorial correspondence search space. This was accomplished using a variety of heuristically set thresholds. Their method solved for the six rigid parameters. Yang et al. (2016) used semantic features from urban scenes for automated TLS co-registration. The point cloud data is segmented into ground and non-ground followed by the extraction of vertical linear features. The vertical features were then triangulated to form a network. Then a hashing table with triangular constraints were used to find matching source and target triangles. The method used various Euclidean distance-based constraints and thresholds which can only be applied when source and target point clouds are of the same scale.

Linear features extracted from point clouds have been used to match Airborne Laser Scanning (ALS) and TLS data (von Hansen et al., 2008). This method sequentially computed the 3D rotation and translation parameters. Rotation was derived via the correlation of line orientation histograms. Afterwards, translation was determined using a 'generate and test' scheme, where the quality of all line correspondence combinations are assessed using the proximity of matching between ALS and TLS line midpoints. Yang et al. (2015) presented an approach for ALS to TLS alignment in urban scenes. They employed a spectral graph correspondence approach for matching building outlines. The graph matching utilized scale-variant geometric constraints such as distances together with several other spatial relations derived from the TLS and ALS building outlines. Urban areas typically contain many other rich descriptive details such as road networks, street furniture and vegetation. Therefore, the method may falter in urban datasets where there is a lack of building structures.

Aiger et al. (2008) developed the '4-Point Congruent Set' (4PCS) method for coarse rigid alignment of point clouds. The approach begins by sampling four-point coplanar tuples from the source point cloud, followed by a search based on an affine ratio to find corresponding four-point tuples in the target point cloud. The best transformation is then selected from multiple candidate transformations formed by the set of matching quad- ruples. There have been several extensions/variations of 4PCS. Theiler et al. (2014) combined 3D keypoints with the 4PCS for the alignment of terrestrial laser scans. In other work, Mellado et al. (2014) developed a speeded up version of 4PCS. In context of full coarse registration (i.e., solving for scale and rigid parameters), Corsini et al. (2013) presented an extension of 4PCS which can handle scale changes between datasets.

2.3. 2D image-based methods

Another active branch of research which addresses the coarse point cloud alignment are image-based approaches. The concept revolves around the utilization of image-based representations of the point cloud data collected from various sensor acquisition systems. We briefly summarize the various types of image-based point cloud representations. One type of image representation can be obtained from optical cameras which are mounted to and synchronised with the laser scanners during point cloud data collection. If the transformation between the camera coordinate system and the laser scanning system is established prior to data collection, then the relative orientation of an image pair can be used to derive