

COMPARISON OF 2D AND 3D APPROACHES FOR THE ALIGNMENT OF UAV AND LIDAR 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
{ravi071, armenc}@yorku.ca

Commission I, ICWG I/II

KEY WORDS: Alignment, Point Clouds, UAV, lidar, Automation, Comparison

ABSTRACT:

The automatic alignment of 3D point clouds acquired or generated from different sensors is a challenging problem. The objective of the alignment is to estimate the 3D similarity transformation parameters, including a global scale factor, 3 rotations and 3 translations. To do so, corresponding anchor features are required in both data sets. There are two main types of alignment: i) Coarse alignment and ii) Refined Alignment. Coarse alignment issues include lack of any prior knowledge of the respective coordinate systems for a source and target point cloud pair and the difficulty to extract and match corresponding control features (e.g., points, lines or planes) co-located on both point cloud pairs to be aligned. With the increasing use of UAVs, there is a need to automatically co-register their generated point cloud-based digital surface models with those from other data acquisition systems such as terrestrial or airborne lidar point clouds. This work presents a comparative study of two independent feature matching techniques for addressing 3D conformal point cloud alignment of UAV and lidar data in different 3D coordinate systems without any prior knowledge of the seven transformation parameters.

1. INTRODUCTION

Co-registration is the process of aligning multiple shapes in a common coordinate system. It is typically applied to overlapping vector datasets, pairs of 2D images or 3D point cloud models. This work concentrates on addressing the latter issue of automated 3D point cloud co-registration. There are two types of alignment problems: *refinement-based* and *coarse* (Figure 1). This work explores two different approaches for coarse alignment. Both methods follow an automated feature matching workflow, including feature detection, feature attribute generation and feature correspondence. However, they are performed in different feature spaces, that is, one approach is 3D based whereas the other approach is 2D based. The 3D based method extracts 3D ‘anchor’ (or ‘key’) points (i.e., distinct points of interest) using local surface curvature cues. Feature description of these anchor points takes into account two local neighbourhood attributes: geodesic surface distance and surface slope. The 2D based method first generates a height image and 2D anchor points are extracted using a scale-space wavelet approach. This is followed by 2D feature description achieved using a combination of the log polar transformation and local Gabor wavelets. Both the 2D and 3D based methods are scale, rotation and translation invariant. Optimal anchor point correspondences are determined by evaluating their descriptor similarity. The matched anchor points in the source and target datasets are then used to estimate the seven conformal transformation parameters.

The primary objective of this work is to evaluate the co-registration performance of each of the two methods on point

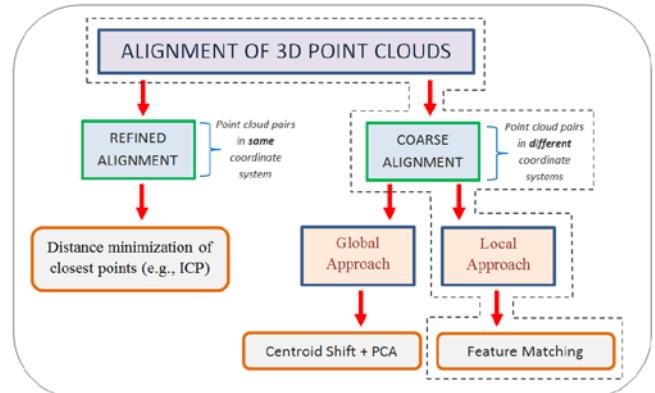


Figure 1. Distinction amongst various 3D point cloud alignment (co-registration) approaches (this work concentrates on the framework marked by the dashed outline).

cloud datasets with varying properties related to the sensor types and the point cloud characteristics. Different sensor types are employed because these introduce additional challenges for the feature correspondence process (Figure 2).

3D point clouds can have varying characteristics and be represented in various ways. They can be represented in 3D or 2D formats such as: i) the default, raw 3D points, or as ii) interpolated, 2D height (or depth) map raster images. As shown in Figure 3, source and target point cloud datasets can also differ in terms of characteristics such as: i) point density (for

* Corresponding author

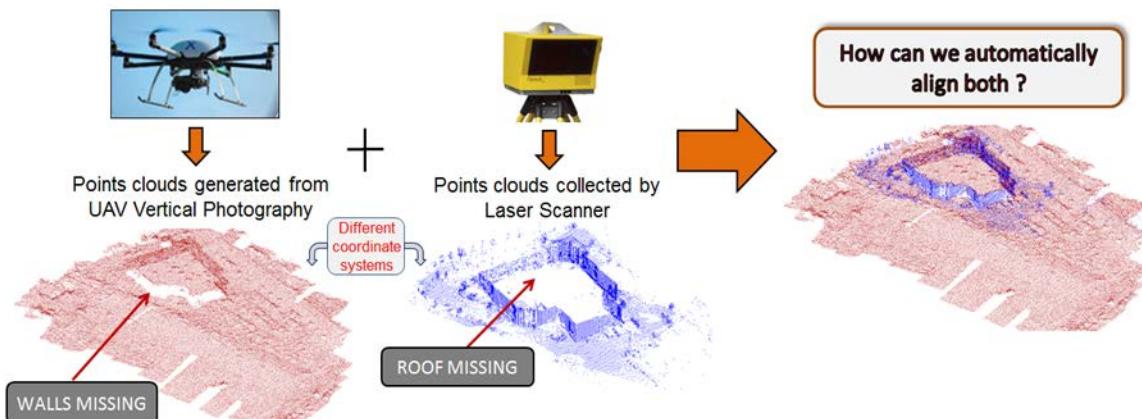


Figure 2. Illustration of the co-registration problem for 3D point cloud datasets from multiple sensors.

example, dense versus sparse point spacing), ii) point distribution (for example, regular, gridded points versus irregular, non-gridded points), and iii) missing point data (i.e., data gaps/holes), possibly caused by occlusions or by different sensor viewpoint perspectives during data acquisition. To handle these different cases (i.e., differences in data representation and characteristics) two novel, independent approaches for the automatic co-registration of point clouds are explored and presented.

the feature matching process is applied to the planimetric, height map projections (i.e., 2D image representation) of the 3D point clouds. For the latter, even though feature matching is performed in the 2D domain, the resulting matched points also have an associated ('Z') or depth component, thereby facilitating 3D to 3D co-registration.

2. RELATED WORKS

Co-registration of 3D point clouds can be achieved using different types of approaches. Alignment can be achieved using descriptor and non-descriptor-based matching workflows. The descriptor-based approaches include the use of the 2D and the raw 3D representations of 3D point cloud datasets. Feature matching can be used to obtain keypoint correspondences required for 3D point cloud co-registration using 2D images (i.e., point clouds represented as intensity images or depth images).

In terms of descriptor-based approaches, (applicable to both the 3D descriptor-based, as well as the 2D image-based categories) there are three main stages involved: i) extraction of keypoints (2D or 3D), ii) formation of keypoint descriptors and iii) correspondence determination based on similar keypoint descriptors. Keypoint detectors can be regarded either as: i) a fixed scale detector, where the user has to define a local neighbourhood around a query point to check if it can be established as a keypoint, or ii) a scale-invariant detector, where the local scale (i.e., local neighbourhood of interest) around a keypoint is automatically defined by the algorithm. This local neighbourhood region is then used to define keypoint attributes (or descriptor).

Common feature detectors and descriptors typically applied for image matching include the SIFT (Lowe, 2004) and SURF (Bay et al., 2008) feature detectors and descriptors. For instance, Weinmann and Jutzi (2011) use both range and reflectance image-based information for determining SIFT correspondences to match terrestrial laser point cloud scans. However, a large quantity of work is also in the area of 3D-based detectors and descriptors for point cloud alignment. 3D keypoint detectors and descriptors can be categorized into either surface geometry-based or volumetric-based algorithms. The difference of these two categories is based on 3D point cloud representation, i.e.,

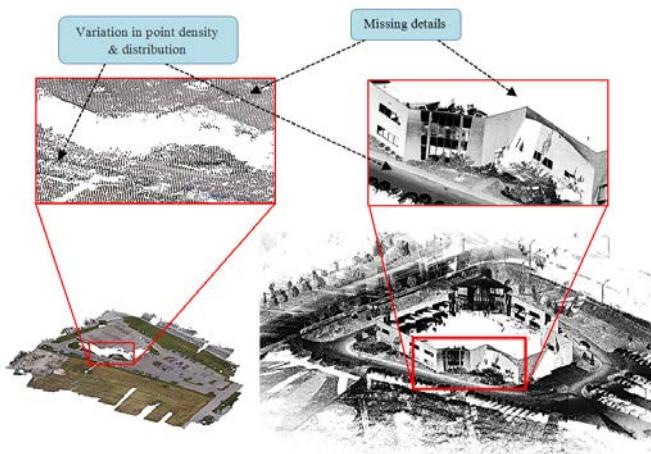


Figure 3. Example of two point cloud datasets from different sensors (left: UAV, right: Mobile laser scanner) with varying point characteristics such as different point density, point distribution and point details.

Although both of the proposed approaches adopt a similar feature matching workflow (Figure 4), their inherent individual components are unique, i.e., the techniques used for keypoint extraction, keypoint descriptor formation, and keypoint matching are different. This stems from the two different ways in which the point clouds can be represented, i.e., either as 3D points or as interpolated, height map raster images. For the first proposed approach, feature matching is performed entirely in the original 3D point cloud spaces, whilst in the second method;

either as raw 3D point clouds (in the case of the geometry-based methods) or as voxelized grids (in the case of volumetric-based methods). Surface geometry-based detectors include Intrinsic Shape Signatures (ISS, Zhong, 2009), mesh-DoG (Zaharescu et al., 2009) and Harris 3D (Sipiran and Bustos, 2011). ISS is a fixed scale detector. ISS uses the ratios of the eigenvalues of the local neighbourhood to determine surface variation. Points with large surface variations are marked as keypoints. The mesh-DoG and Harris 3D detectors operate on mesh representations of the point clouds.

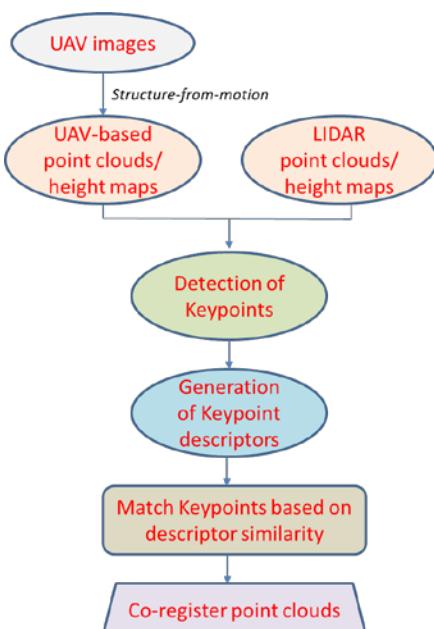


Figure 4. Illustration of the feature matching pipeline used for 3D point cloud alignment. The input data for this workflow can be the raw 3D points or height map image representations of the point clouds.

The mesh-DoG is scale-invariant, using a ‘Difference-of-Gaussian’ approach (i.e., smoothing the data with Gaussian kernels of differing standard deviations) for scale-space representation. For mesh-DoG, the ratios of eigenvalues from the local mesh neighbourhood are used for keypoint definition. Harris 3D is not a scale-invariant detector. It fits a local surface quadratic patch to the point data and computes the so-called ‘Harris-response’ (Harris and Stephens, 1988) for each mesh vertex. Query vertices with large responses are classified as keypoints.

There has also been various surface geometry 3D point cloud descriptors developed over the years. Some of these include Spin Images (Johnson and Hebert, 1999), Fast Point Feature Histograms (FPFH) (Rusu et al., 2009) and Signature of Histograms of Orientations (SHOT) (Tombari et al., 2010). For Spin Images, every point within the local keypoint neighbourhood are assigned two coordinates, α and β ; α is the distance from the keypoint to the projection of the neighbourhood point on the local surface tangent plane. β is the distance from the neighbourhood point to the local tangent plane. For every point in the local neighbourhood, these pair of coordinates is accumulated into a 2D array, thus forming the descriptor. FPFH is a histogram-based descriptor which bins three angular attributes defined by the relation between every

neighbourhood point and the keypoint. SHOT is also a histogram-based descriptor which defines a spherical neighbourhood around the keypoint. This spherical neighbourhood is then partitioned into spherical grid sectors. For each grid sector, the angles between the normals at the neighbouring points and the normal at the keypoint are accumulated into a local histogram. The local histograms of all grid sectors are then concatenated to form the SHOT descriptor.

There are also volume-based methods which utilize 3D voxel representations instead of direct point cloud data for keypoint detection and keypoint description. These include a 3D extension of the 2D SIFT method (Lowe, 2004). The 3D-SIFT keypoint detector (Filipe and Alexandre, 2014) is scale-invariant and utilizes a ‘Difference-of-Gaussian’ scale-space approach, where a series of downsampling / smoothing is applied to the point data to determine keypoints and their respective local scales. There is also volumetric 3D-SIFT descriptors (Flitton et al., 2013) which are based on gradient histogram attributes of the voxel grids.

Non-descriptor 3D point cloud co-registration methods include methods such as principal component analysis (PCA), which can be usefully applied given that there is complete overlapping coverage for the pairwise point clouds to be aligned. Recently, the 4-Points Congruent Sets (4PCS) (Aiger et al., 2008) has also been popularly applied for 3D point cloud surface alignment. It utilizes geometric, affine-invariance ratios to extract co-planar, congruent points on the source and target shape which are then used for aligning the datasets.

3. METHOD

We compare two methods for 3D point cloud alignment. The first approach (Persad and Armenakis, 2017a) uses the raw 3D point clouds for 3D feature matching. The second approach (Persad and Armenakis, 2017b) uses height map projections of the point clouds for 2D feature matching. Both of the methods are invariant to scale, 3D rotation and 3D translation, i.e., they can be applied to co-register point cloud datasets in different 3D conformal coordinate systems. The methods are briefly summarized in this section.

3.1 3D-based feature matching

This method (Persad and Armenakis, 2017a) firstly detects 3D scale-invariant keypoints on point cloud surfaces. The detector follows a scale-space approach to estimate the local keypoint scales to be used for defining the descriptor regions. The scale-space mechanism is based on the definition on multiple, concentric circular neighbourhoods around a keypoint candidate (Figure 5). For each concentric region, the surface curvature for the point is estimated using the local point clouds defined within the region. This procedure is repeated for a series of normalized radii values. A search is then performed to detect distinct maxima based on the set of curvature values. For the points with a strong maxima response, a further check is then applied to compare its curvature strength relative to its neighbouring point clouds. A keypoint is detected if the strength of the candidate keypoint is higher than its neighbours. Adaptive non-maxima suppression is also used to filter weak keypoints.

The radii maxima determined during the keypoint stage is then used to define the region for generating the keypoint’s 3D

descriptor. Local surface morphological attributes are used to populate the keypoint's descriptor space. In particular, the radial geodesic distance projecting from the keypoint to its neighbours, as well as the local surface slope around each of the neighbouring points are used. 3D rotation and translation invariance of the descriptor is achieved through the definition of a local reference frame. The keypoint neighbourhood is then transformed to this local system. Bipartite graph matching and a RANSAC-based outlier removal algorithm are then used to find one-to-one descriptor correspondence.

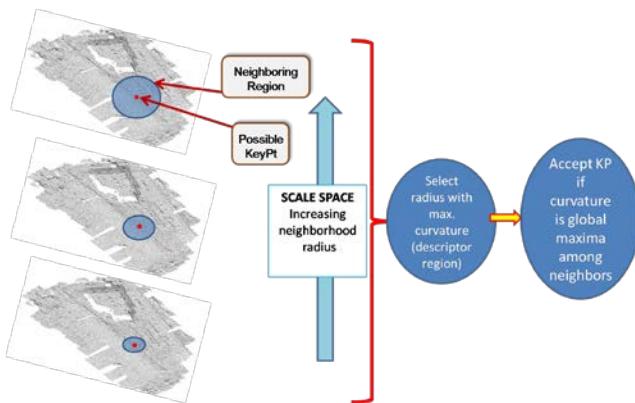


Figure 5. Concept of the 3D scale-invariant keypoint detector.

3.2 Height map-based feature matching

This approach (Persad and Armenakis, 2017b) projects the 3D points onto a 2D image plane where raster-based interpolation is then applied. The resultant height map images are then subjected to a multi-scale wavelet-based keypoint detector. The detector uses a complex wavelet transform (Kingsbury, 1998) to identify salient keypoints across scale-space.

Afterwards, 2D keypoint descriptors are computed. Compared to the 3D-based approach, the scale-invariance is achieved during the descriptor phase and not provided by the wavelet-detector. Local log-polar sampled-gridding around the keypoint and mapping to log-polar space modelled any scale and rotation differences between source and target height maps. These differences are reflected as cyclic shifts in the log-polar descriptor space. For a fully-invariant representation, 'Rapid-transform' (Reitboeck and Brody, 1969) was applied. Finally, bi-directional nearest neighbour matching in descriptor space and RANSAC-based outlier removal is employed for correspondence determination.

4. EXPERIMENTS & RESULTS

4.1 Study area and datasets

Two datasets were evaluated for co-registration. Dataset 1 (shown in Figure 3) consists of UAV and mobile laser scanning point clouds. The UAV point cloud was obtained using Structure-from-Motion (SfM) from vertical imagery captured by a 19mm camera mounted on a Geo-X8000 UAV. Dataset 2 (Figure 6) consists of UAV and airborne laser scanning point clouds. In the second case, the UAV platform was an Aeryon Scout and the photogrammetric point cloud was generated from oblique and nadir-looking imagery via SfM.

4.2 Results

Tables 1 and 2 show results achieved by the two approaches in terms of the number of determined correspondences, as well as alignment accuracy. Reference values were estimated by manually selecting 8 point correspondences and computing the parameters of the 3D conformal transformation. The alignment accuracy is reflected by Δs (which is the absolute difference in scale between the automatically estimated and the reference values), mean rotation error (MRE, which is the average absolute error of the 3 rotation angles) and mean translation error (MTE, which is the average absolute error of the 3 translations).

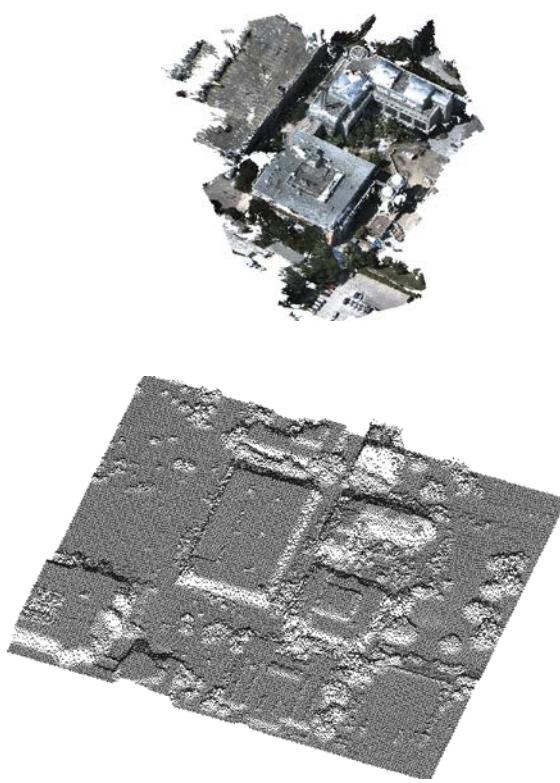


Figure 6. Illustration of the second dataset used for the co-registration experiment. Top: Photogrammetric point clouds generated by SfM from UAV. Bottom: Airborne laser scanning point clouds.

The 3D-based method did not recover any point matches for either case. This was due to the difference in point characteristics between the UAV and lidar datasets, including point density and point distribution. For example, in Dataset 2, the aerial lidar data had a lower point density than the UAV point clouds. The UAV point clouds were irregularly spaced and the aerial lidar was regularly spaced. Additionally, due to the different perspectives of the respective sensor during data acquisition, there were differences in point details (i.e., holes/gaps). For instance, in Dataset 1, the UAV point cloud included roof and wall structure details; while the mobile lidar data does not have any roof information. This resulted in unwanted dissimilarities between 3D descriptors at corresponding keypoint locations on the UAV and lidar datasets. The effect of missing points, and diverseness in point

characteristics were minimized when matching the height-map representations of the point clouds.

Table 1. Results for 3D-based alignment

Error measure	Dataset 1	Dataset 2
# source KPs/target KPs	363/641	330/552
# correct matches	0	0
Δs	-	-
MRE ($^{\circ}$)	-	-
MTE (m)	-	-

Table 2. Results for Height-map based alignment

Error measure	Dataset 1	Dataset 2
# source KPs/target KPs	148/189	215/314
# correct matches	8	11
Δs	0.080	0.010
MRE ($^{\circ}$)	0.207	0.300
MTE (m)	0.090	0.403

5. CONCLUSIONS AND OUTLOOK

Two different 3D point cloud co-registration methods have been evaluated. Experimental analysis showed that the selection of using one co-registration method instead of the other depends strongly on the characteristics of the point cloud dataset. The height map-based method is a more ‘generalized’ 3D point cloud co-registration approach. That is, it is robust to datasets with different point densities and different point distributions since it provides a continuous form of coverage. Furthermore, the height map-based approach is able to co-register multi-sensor datasets where there are missing point clouds (i.e., holes/gaps) due to data being collected from different viewing perspectives (e.g., Aerial and terrestrial point clouds). Future work will investigate approaches to improve the various components of the 3D-based co-registration approach, in an effort to remain robust to changes due to point cloud characteristics and sensor viewing perspectives.

ACKNOWLEDGEMENTS

We wish to thank First Base Solutions, Prof. James Elder, Mike Leslar and Teledyne Optech for providing the data for this work. This work is financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and York University.

REFERENCES

- Aiger, D., Mitra, N.J. and Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface registration. *ACM Transactions on Graphics (TOG)*, 27(3), p.85.
- Bay, H., Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. 2008. Speeded-up robust features (SURF). Computer vision and image understanding 110, no 3: 346-359.
- Filipe, S. and Alexandre, L.A., 2014. A comparative evaluation of 3D keypoint detectors in a RGB-D object dataset.

In Proceedings of *Computer Vision Theory and Applications (VISAPP)*, 2014, Vol. 1, pp. 476-483.

Flitton, G., Breckon, T.P. and Megherbi, N., 2013. A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery. *Pattern Recognition*, 46(9), pp. 2420-2436.

Kingsbury, N., 1998. The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement. In *IEEE Signal Processing Conference (EUSIPCO 1998)*, 9th European (pp. 1-4).

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60 (2), 91–110.

Persad, R.A. and Armenakis, C., 2017a. Automatic 3D Surface Co-Registration Using Keypoint Matching. *Photogrammetric Engineering & Remote Sensing*, 83(2), pp.137-151.

Persad, R.A. and Armenakis, C., 2017b. Automatic co-registration of 3D multi-sensor point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, pp.162-186.

Reitboeck, H. and Brody, T.P., 1969. A transformation with invariance under cyclic permutation for applications in pattern recognition. *Information and Control*, 15(2), pp.130-154.

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

Sipiran, I. and Bustos, B., 2011. Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. *The Visual Computer*, 27(11), pp. 963-976.

Tombari, F., Salti, S. and Di Stefano, L., 2010. Unique signatures of histograms for local surface description. *Computer vision-ECCV 2010*, pp. 356-369.

Weinmann, M. and Jutzi, B., 2011. Fully automatic image-based registration of unorganized TLS data. *Int Arch Photogramm Remote Sens Spat Inf Sci*, 38(5).

Zaharescu, A., E. Boyer, K. Varanasi, and R. Horaud, 2009. Surface feature detection and description with applications to mesh matching, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2009*, pp. 373–380.

Zhong, Y. 2009. Intrinsic shape signatures: A shape descriptor for 3D object recognition, *Proceedings of the International Conference on Computer Vision Workshops*, pp. 689–696.