ROBUST CYLINDER FITTING
IN THREE-DIMENSIONAL POINT CLOUD DATA

Abdul Nurunnabi [a,*] , Yukio Sadahiro[b], Roderik Lindenbergh, [c]

[a,b] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan
[c] Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands
[a,b]{a.nurunnabi, sada}@csis.u-tokyo.ac.jp, [c] r.c.lindenbergh@tudelft.nl

Commission II, WG II/3

KEY WORDS: Feature Extraction, Geometric Shape, Laser Scanning, Object Recognition, Pole Modelling, Robust PCA, Robust Regression, Surface Fitting

ABSTRACT:

This paper investigates the problems of cylinder fitting in laser scanning three-dimensional Point Cloud Data (PCD). Most existing methods require full cylinder data, do not study the presence of outliers, and are not statistically robust. But especially mobile laser scanning often has incomplete data, as street poles for example are only scanned from the road. Moreover, existence of outliers is common. Outliers may occur as random or systematic errors, and may be scattered and/or clustered. In this paper, we present a statistically robust cylinder fitting algorithm for PCD that combines Robust Principal Component Analysis (RPCA) with robust regression. Robust principal components as obtained by RPCA allow estimating cylinder directions more accurately, and an existing efficient circle fitting algorithm following robust regression principles, properly fit cylinder. We demonstrate the performance of the proposed method on artificial and real PCD. Results show that the proposed method provides more accurate and robust results: (i) in the presence of noise and high percentage of outliers, (ii) for incomplete as well as complete data, (iii) for small and large number of points, and (iv) for different sizes of radius. On 1000 simulated quarter cylinders of 1m radius with 10% outliers a PCA based method fit cylinders with a radius of on average 3.63meter (m); the proposed method on the other hand fit cylinders of on average 1.02 m radius. The algorithm has potential in applications such as fitting cylindrical (e.g., light and traffic) poles, diameter at breast height estimation for trees, and building and bridge information modelling.

1. INTRODUCTION

Laser scanning has been recognized as an effective and high speed survey tool, for direct acquisition of dense three-Dimensional (3D) spatial data called point clouds. Point clouds achieve large metric precision at moderate costs. Due to the complexity and large volume (i.e., big data) of Point Cloud Data (PCD), extracting meaningful geometric shapes has an ever-growing demand. Cylindrical features are one of the most encountered geometric primitives in point clouds when collected in industrial sites, urban areas, street corridors, but also forest. Cylinder fitting is a fundamental task in computer graphics, computer vision, image analysis, reverse engineering, pattern recognition, photogrammetry and remote sensing for various applications like feature extraction (De Guevara et al., 2011), surface reconstruction (Schnabel et al., 2007), archaeological documentation, street furniture management, building information modelling (Kwon et al., 2004), as-built modelling (Ahmed et al., 2014), machine tools quality control, industrial plant settings (Su et al., 2016), tunnel monitoring (Gosliga et al., 2006), and forest inventory (Lalonde et al., 2006), and therefore has been studied rigorously.

PCD are usually partial, unordered, have irregular point density, contain various complex structures, and may be noisy. As well, the presence of outliers that do not follow the outline of the majority of points of interest is common for many reasons such as systematic biases or sensor errors, multipath reflections, occlusions, unexpected moving objects, environmental disorder like snow, rain, dust, and random noise that may appear as off-surface points. Pseudo-outliers that are created by multiple and unorganized structures mainly appeared as clustered outliers (cf., Nurunnabi et al., 2015). PCD especially from Mobile Laser Scanning (MLS) and Aerial Laser Scanning (ALS) are mostly incomplete and sometimes locally missing because the corresponding scan systems obtain PCD from a particular point of view. The same holds often when using Terrestrial Laser Scanning (TLS); at many archaeological sites, power plants and in industrial settings it is not always possible to scan around the objects of interest because of the scene complexities do not allow the required physical positioning of the scanners. Therefore, robust cylinder fitting for incomplete PCD with outliers has great importance and but is also challenging.

Indeed, much research has been carried for cylinder fitting in 3D data (Shakarji, 1998; Marshal et al., 2001; Rabbani and Heuvel, 2005; Schnabel et al., 2007; Mendez et al., 2014; Tran et al., 2015). Based on data and applications, developments have been made into two main classes; some algorithms work directly on raw data without significant pre-processing, others take pre-processed (e.g., classification and segmentation) data so that every part is well-approximated by an appropriate geometric shape. RANdom SAmple Consensus (RANSAC; Fischler and Bolles, 1981) and the Hough Transform (HT; Duda and Hart, 1972) are two well-known voting techniques used for shape detection and fitting in raw 2/3D data (Schnabel et al., 2007). Although the HT and RANSAC can tackle the influence of outliers, many people are not happy to use them because of time and/or space complexity. Nurunnabi et al. (2014, 1015) showed that RANSAC is influenced by the well-known masking effect (cf., Rousseeuw and Leroy, 2003).

---

* Corresponding author

Chaperon and Goulette (2001) proposed a two-step approach; in the first step RANSAC is used on a Gaussian sphere to find the cylinder's direction while a second step finds the cylinder's position and size. Beder and Förstner (2006) used RANSAC to get a direct solution on the original point cloud for estimating cylinder parameters. Schnabel et al. (2007) used the RANSAC algorithm for geometric shape (e.g., planes and cylinders) detection. The authors mentioned that their method can be applied to a large variety of data and is extremely robust, but does not find shape proxies for every part of the surface. Vosselman et al. (2004) introduced a HT based cylinder detection algorithm that uses normals of the data points to get the cylinder direction. Rabbani and Heuvel (2005) developed two sequential steps of low dimensional HT for detecting cylinders in PCD. Many authors (e.g., Tarsha-Kurdi et al., 2007; Deschaud and Goulette, 2010) made comparisons between HT and RANSAC, and showed that HT is less efficient than RANSAC in terms of computation time when a model is being fitted to a large data set, while, in addition, HT is sensitive to the choice of segmentation parameters for segment based modelling. Over the years many revised versions of RANSAC have been developed. This paper considers RANSAC (Fischler and Bolles, 1981) for comparison as it is well-known, and many people (cf., Wang et al., 2016) are using it as their first choice, considering its high-quality results, robustness and generality.

The methods that do pre-processing before cylinder fitting can be grouped into non-linear and linear least squares approaches. Linear least squares methods are based on a linearization of the Euclidean distance. On the other hand, non-linear methods minimize the geometric distance of the points to the surface of a cylinder (Marshall et al., 2001). Franaszek (2012) studied the variability of cylinder parameters for non-linear fitting. Rahayem et al. (2012) fit cylinders followed by ellipse fitting for data without outliers. Kwon et al. (2004) developed a PCA based cylinder fitting algorithm without using the normals of the data. Lalonde et al. (2006) worked in a similar way (Kwon et al., 2004), and developed two algorithms based on 2D projection and 3D fitting for tree trunk segmentation. They showed that results from the PCA based 2D projection technique approximately 25% better than the results from 3D fitting. The authors (Kwon et al., 2004; Lalonde et al., 2006) claimed their methods are computationally efficient but both of them did not consider yet data with gross or pseudo-outliers, which is one of the main concerns in this paper.

Existing cylinder fitting methods mainly focused on full data without outliers. This paper concentrates on incomplete 3D PCD in the presence of outliers. Fig. 1 shows MLS signpost data, which is incomplete and has many outlying points in the red ellipses [Fig.1(d)] which become visible if we look horizontally (top view) through the post. This paper proposes an algorithm based on the PCA approach (Lalonde et al., 2006) but it notably uses a robust version of PCA to get robust principal components. Those will be consecutively used to get cylinder directions and 2D points by projecting onto two PC based plane. The new method couples robust regression with circle fitting to get reliable estimates of cylinder parameters. It fits a cylinder model to pre-processed partial cylindrical data. Pre-processing could for example consist of a segmentation step (Nurunnabi et al., 2016). The method is statistically robust and consistent, produces reliable results in the presence of a high percentage of clustered outliers for incomplete as well as full data.

The remaining of this paper is arranged as follows. The new algorithm is described in detail in Section 2. Section 2.1 discusses technical issues and principles used in the developed

algorithm. The algorithm itself is described in Section 2.2. Detailed experiments on artificial and laser scanning data are evaluated in Section 3. Section 4 concludes the paper.
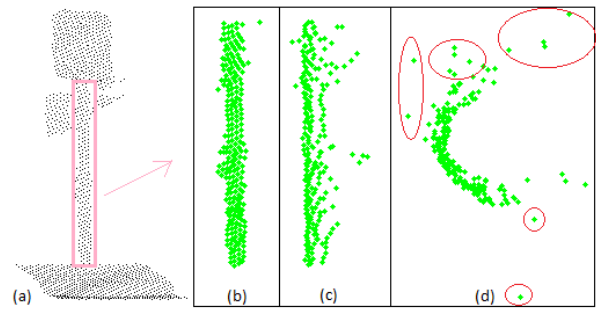

Figure 1. (a) Mobile laser scanned signpost; incomplete point cloud data, (b) a segment (cylindrical part; front view) of the data, (c) cylindrical data; side view, and (d) cylindrical data; top view, outliers are highlighted by red ellipses

2. PROPOSED ALGORITHM

The robust cylinder fitting algorithm proposed in this section uses robust statistical approaches (robust PCA and robust regression) and an existing circle fitting method. We briefly include technical issues and principles of the existing tools, and then describe the new algorithm in detail.

2.1 Robust Statistical Approaches and Algebraic Circle Fitting

2.1.1 Robust Principal Component Analysis: Principal Component Analysis (PCA) presents a set of orthogonal and uncorrelated variables, which are linear combinations of the mean centered original variables, called Principal Components (PCs). For a 3D point cloud, PCs ($v_i; i = 2,1,0$) correspond with the eigenvalues ($\lambda_i; i = 2,1,0;$ usually $\lambda_2 \geq \lambda_1 \geq \lambda_0$) of the covariance matrix $\Sigma_{3\times3} = 1/n \sum_{i=1}^{k}(p_i - \bar{p})(p_i - \bar{p})^T$, where $p_i(x_i, y_i, z_i)$ is the i-th point in the point cloud. The eigenvalues express the variability of the data along the corresponding eigenvectors. Hence, the first PC (v_2) shows the highest ($\lambda_2\%$) variability of the data. Although PCA has been used successfully in point cloud processing, the results from PCA are highly influenced by outliers and produce misleading results (Nurunnabi et al., 2014, 2015, 2016).

To overcome the sensitivity towards outliers, many Robust versions of PCA (RPCA) have been developed over the years (e.g., Hubert et al., 2005). We use RPCA (Hubert et al., 2005), which is good for lower dimensional (3D) point cloud processing (Nurunnabi et al., 2014). It combines the so called projection-pursuit approach with robust covariance estimation. In RPCA, for each point p_i, a measure of outlyingness is defined as

$$w_i = argmax_v \frac{|p_i v^T - \mu_{\text{FMCD}}(p_i v^T)|}{\Sigma_{\text{FMCD}}(p_i v^T)}, \qquad (1)$$

which is computed by projecting the data onto many univariate directions. For every direction, the Fast Minimum Covariance Determinant (FMCD; Rousseeuw and van Driessen, 1999) based robust centre μ_{FMCD} and covariance matrix Σ_{FMCD} of the projected points $p_i v^T$ are computed in Eq. (1). Next, a majority share ($h > n/2$) of points having minimum values of w_i are used to construct a robust covariance matrix that is used later for PCA and to get robust PCs. Hubert et al. (2005) claimed that the resultant robust PCs are location and orthogonal invariant, and their method produces accurate estimates for data without outliers and more robust estimates for data with outliers.

2.1.2 Robust LTS Regression: Consider the standard regression model

$$Y = X\beta + \varepsilon, \tag{2}$$

where Y is the response variable, X is an $n \times k$ ($k = m + 1$) matrix of explanatory variables including one constant column of 1's, β is a $k \times 1$ vector of parameters, and ε is an $n \times 1$ random error vector. To estimate the parameters, the well-known Least Squares (LS) method minimizes the sum of the squared residuals, i.e.,

$$minimze_{\hat{\beta}} \sum_{i=1}^{n} e_i^2, \tag{3}$$

where $e_i = Y_i - \hat{Y}_i$, $\hat{Y}_i = X_i^T \hat{\beta}$, and $\hat{\beta}$ is the vector of estimated parameters. Although LS has many good statistical properties, it is sensitive to outliers and may produce deceptive results. To minimize the influence of outliers, many robust and diagnostic regression methods have been developed (cf., Rousseeuw and Leroy, 2003). Many authors replace the squared residuals in Eq. (3) by some functions of residuals such as M-estimators (cf., Wang and Suter, 2003), which are robust, but they are vulnerable to leverage in terms of X points and do not achieve a high Breakdown Point (BP; the level of outlier tolerance). Rousseeuw (1984) proposed two well-known high BP robust regression methods: Least Median of Squares (LMS) and Least Trimmed Squares (LTS). In LMS, the median of squared residuals is minimized, defined as

$$minimize_{\hat{\beta}} \ median_i \ e_i^2. \tag{4}$$

LTS minimizes the trimmed sum of squared residuals, i.e., this algorithm minimizes the sum of the h ($h \geq n/2$) lowest squared residuals, i.e.,

$$minimize_{\hat{\beta}} \sum_{i=1}^{h} e_i^2. \tag{5}$$

LMS and LTS both achieve 50% BP, when h is close to $n/2$. For LMS the parameters are solved by non-linear minimization, therefore an exact solution cannot be obtained. LMS is computationally intensive with low convergence rate. We adapt LTS in our algorithm, because LTS is much faster, asymptotically normal, and it has better statistical efficiency (cf., Rousseeuw and Leroy, 2003; Wang and Suter, 2003).

2.1.3 Hyper Circle Fit: Circle fitting methods consist of mainly two groups: (i) geometric, and (ii) algebraic. The first one minimizes the geometric distances, and the second minimizes algebraic functions. The geometric fits are computationally intensive, and the algebraic fits are faster because they are non-iterative unlike geometric fits. We emphasize on faster methods as we are dealing with large PCD. Algebraic fits minimize the standard function:

$$\mathcal{F} = \sum[A(x^2 + y^2) + Bx + Cy + D]^2, \tag{6}$$

where x, y are the points for circle fitting in 2D space; A, B, C and D are the circle parameters with $A \neq 0$. Eq. (6) can be rewritten as

$$\mathcal{F} = \sum(z_i + Bx_i + Cy_i + D)^2, \tag{7}$$

where $A = 1$ and $z_i = x_i^2 + y_i^2$; $i = 1,2, \dots, n$. This is a system of linear equations which can be solved w.r.t. B, C and D. If $a = -B/2A$, $b = -C/2A$ and $r = \sqrt{B^2 + C^2 - 4AD}/2A$, then the algebraic parameters can be changed to the geometric parameters: centre (a, b) and radius r for the circle that minimizes sum of the squared residuals

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [\sqrt{(x_i - a)^2 + (y_i - b)^2} - r]^2, \tag{8}$$

where e_i is the i-th residual. Based on different parameter constraints, many algebraic methods have been developed to solve Eq. (7). One of the most efficient is developed by Al-Sharadqah and Chernov (2009), called 'Hyper' that minimizes Eq. (7) w.r.t. $\boldsymbol{A}^T N_H \boldsymbol{A} = 1$, where $\boldsymbol{A} = (A, B, C, D)$ and $N_H = 2N_T - N_P$, where N_T and N_P are the constraints $4A^2\bar{z} + 4AB\bar{x} + 4AC\bar{y} + B^2 + C^2 = 1$ and $B^2 + C^2 - 4AD = 1$, respectively. The authors (Al-Sharadqah and Chernov, 2009) made comparisons with many of the algebraic fits, and pointed out that their fit outperforms other algebraic fits with less error. We adapted the Hyper method because, among other benefits, this method produces results that are invariant under rotation and translation (cf., Al-Sharadqah and Chernov, 2009).

2.2 Algorithm Implementation

The cylinder fitting framework proposed in this section incorporates RPCA to estimate robust PCs that find cylinder position and direction more accurately. 3D cylinder points are projected onto the plane fixed by the two first PCs. Ideally, the projected points outline a circle. To estimate the radius of the cylinder we fit a circle to these projected 2D points. We want to use an efficient robust circle fitting method, because the projected points in practice may only be approximately circular and may contain outliers as we did not remove outliers using RPCA. Existing methods however are sensitive to outliers, that means they are statistically non-robust. For robust circle fitting we couple the LTS regression-principle with the Hyper circle fitting method. Using LTS regression, the new algorithm trims $(n - h)$ points having the largest sum of the squared residuals, and fits the circle with the most consistent h ($h \geq n/2$) points that are most close to the circular arc. To estimate the four following cylinder parameters: orientation (O), length (L), centre (C), and radius (R), we perform the following steps that together build the new algorithm.

(i) Apply RPCA to the cylindrical 3D point cloud P, and find three robust PCs.
(ii) As the first PC v_2 corresponds to the highest variability of P, which is the cylinder's length (assuming that its length is larger than its width), we use the first PC to get the orientation of the fitted cylinder, i.e., $O = v_2$. Projection of the cylinder points onto the first PC is the one dimensional representation of the cylinder.
(iii) The length of the cylinder is calculated as the distance between the maximum and minimum values of the projected inlier points onto the first PC. It is defined as

$$L = max_i(p_i^T v_2) - min_i(p_i^T v_2). \tag{9}$$

(iv) The second and third PCs form an orthogonal basis to generate a plane. Project the 3D cylinder points $p_i(x_i, y_i, z_i) \in P$ onto the plane defined by these two PCs to produce 2D circular points $q_i(x_i, y_i) \in Q$ that represent the circular arc for the cylinder ends of the cylinder with equal radii. The i-th 2D point $q_i(x_i, y_i)$ is defined as

$$q_i(x_i, y_i) = (v_o \ v_1)^T (p_i - \bar{p}). \tag{10}$$

The Q is in general not free from outliers as it contains all the projected points from P. We use the Hyper method for circle fitting, but, it is as demonstrated in Fig. 2(b), the Hyper method does not tolerate outlier effects, even in case of full circle data. To reduce outlier effects on circle fitting, we accommodate the LTS regression principle, i.e., consider the majority h ($h > n/2$)

of points that have ordered minimum squared residuals according to the Hyper method. We fix $h = \lceil n/2 \rceil$, which allows us to achieve circle parameters with a high BP of around 50%. To get the best h subset (points that are closest to the circle arc) of points, we iterate the following Steps (a–c) a number of times I_n, which is calculated by the Monte Carlo Probabilistic approach (cf., Fischler and Bolles, 1981). I_n depends on the probability p_r of getting at least one outlier-free minimal subset of h_0 points from data contaminated with an ϵ portion (assume 50% in our algorithm) of outliers. p_r and I_n are related as

$$I_n = \frac{\log(1-p_r)}{\log(1-(1-\epsilon)^{h_0})}. \tag{11}$$

A minimal subset for circle fitting consists of three points, and we fix a large probability $p_r = 0.999$ to get accurate results.
(a) Fit a circle from a minimal h_0-subset, calculate its centre (a_0, b_0), radius r_0 and squared residuals e_i^2 as defined in Eq. (8).
(b) Sort the squared residuals; pick the h points having minimum squared residuals.
(c) Fit a circle with the h points from Step (b); calculate and sort the squared residuals in increasing order, and sum the squared residuals for the first h sorted squared residuals.
Repeat the process I_n times, and chose the h-subset for which the sum (for the h sorted squared residuals) in Step (c) is minimal. Fit a circle to the final h subset, and estimate the circle parameters. The estimated circle centre and radius are the required centre C and radius R for the cylinder. Hence, we create the robust fitted cylinder with parameters: O, L, C, and R.
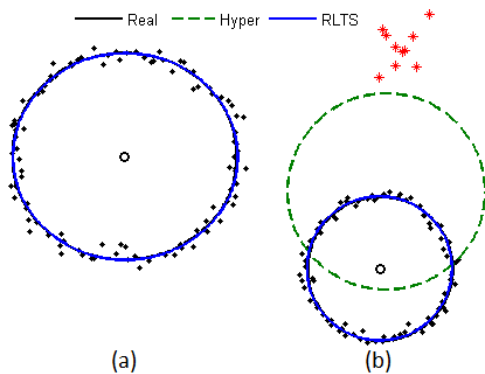

Figure 2. Influence of clustered outliers on circle fitting; (a) data without outliers, (b) data with outliers (red points)

3. EXPERIMENTS AND EVALUATION

The proposed cylinder fitting algorithm is demonstrated and evaluated on several artificial and real point clouds. We evaluate the performance of the new algorithm in terms of data completeness, robustness, consistency, and size of the cylinder's radius, and compare results with three existing methods: (i) LS method, (ii) a PCA based method (Lalonde et al., 2006), in the following referred to as Lalonde, and (c) RANSAC.

3.1 Measures of Evaluation

To evaluate the parameters C, R, L, and O of the fitted cylinder; we calculate the Average Distance between the real cylinder centres and the fitted cylinder centres $AD(\hat{C})$, the estimated Average radius $A(\hat{R})$, the Average length $A(\hat{L})$, and the Average bias orientation $A(\theta)$ for a number of samples. These measures are defined as

$$\left.\begin{array}{l} AD(\hat{C}) = \frac{1}{m}\sum_i^m |C_i - \hat{C}_i|, \\ A(\hat{R}) = \frac{1}{m}\sum_i^m (\hat{R}_i), \\ A(\hat{L}) = \frac{1}{m}\sum_i^m (\hat{L}_i) \text{ and} \\ A(\theta) = \frac{1}{m}\sum_i^m (\theta_i) \end{array}\right\}, \tag{12}$$

where $\hat{C}_i$, $\hat{R}_i$, $\hat{L}_i$ and θ_i are the estimates of centre, radius, length and bias orientation θ [see Fig. 3] for the i-th sample, respectively, and m is the number of samples (datasets). Bias orientation is defined as the angular deviation between the axis of the real and the fitted cylinder, i.e.,

$$\theta = \arccos|v_2^T \hat{v}_2|, \tag{13}$$

where v_2 and $\hat{v}_2$ are the real and the estimated axes of the cylinder, respectively. We give priority on the accuracy of $AD(\hat{C})$ and $A(\theta)$ above $A(\hat{R})$ and $A(\hat{L})$ since the fitted cylinder should be in the right position and fair to the majority of points. It may be the case that the estimated radius and length are equal to the real radius and length respectively, but that the cylinder is fitted somewhere else, because the presence of similar multiple shapes is common in PCD.
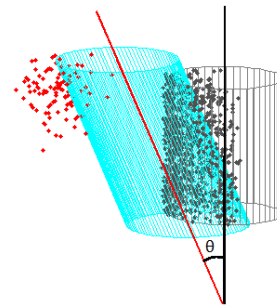

Figure 3. Illustration of bias orientation θ, i.e., angle between real cylinder (grey) axis and fitted cylinder (cyan) axis. Black dots are cylinder points and red points are outliers

3.2 Artificial Datasets

We simulate datasets of a straight cylinder with circular ends of 1000 points, radius R=1meter (m) and length L=10m with axes' end points (0.0, 0.0, 0.0) and (0.0, 0.0, 10.0). Variations among the points are created by adding random noise of Standard Deviation (StD) 0.15 in all (x, y and z) directions. As said, the main focus of this paper is cylinder fitting for incomplete data in the presence of clustered outliers. Clustered outliers are generated with mean (-3.0, 0.0, 9.0) and StD (0.3, 0.3, 1.5). For example, a simulated half cylinder with 10% clustered outliers (red points) is shown in Fig. 3. Although different positions of the outliers w.r.t. the partial cylinder data may result in different outcomes of the methods, it is reasonable that we have to consider outliers somewhere else apart from the cylinder points.

3.2.1 Fitting Cylinders to Complete and Incomplete Data: We check the performance of the cylinder fitting algorithms for incomplete as well as complete data in the presence of outliers. We simulate cylinders including 10% clustered outliers as described above and shown in Fig. 3. To get statistically representative results; we generate 1000 datasets of full and partial cylinders. As partial cylinders we consider cylinders where only half or one quarter of the cylinder is sampled by points in the circular direction. We fit cylinders and calculate the performance measures: $AD(\hat{C})$, $A(\hat{R})$, $A(\hat{L})$ and $A(\theta)$ by LS, Lalonde, RANSAC, and RLTS methods. Results, in Table 1, show that LS and Lalonde perform almost similar for full data. Although most of the cases LS is better than Lalonde for

Methods	Full cylinder				Half cylinder				Quarter cylinder			
	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$
LS	0.55	1.17	13.20	8.78	1.85	1.34	12.79	1.16	3.76	2.87	12.73	1.48
Lalonde	0.75	1.31	13.16	7.64	1.41	1.10	12.98	5.64	4.50	3.63	12.99	5.87
RANSAC	0.40	1.01	12.80	0.96	0.42	1.00	12.80	0.98	0.69	1.19	12.77	1.40
RLTS	0.08	1.00	10.06	0.25	0.09	1.00	10.07	0.31	0.12	1.02	10.07	0.36

Table 1. Evaluation of estimated parameters for full, half and quarter cylinders
[$AD(\hat{C})$ = Average Distance between the real cylinder centres and the fitted cylinder centres, $A(\hat{R})$ = estimated Average radius,
$A(\hat{L})$ = estimated Average length, and $A(\theta)$ = Average bias orientation]

incomplete data, both of them are not acceptable for half and quarter cylinders. They produce remarkably larger radii and lengths than the real cylinder. For quarter cylinders, Lalonde fits a cylinder of on average 3.63 times larger radius than the real radius of 1m. Moreover, the fitted cylinder is 4.50m away from the real cylinder with bias orientation 5.87°. RANSAC performs comparatively better than LS and Lalonde, but significantly worse than the proposed method. For all the cases, RLTS produces cylinders with almost the same radii and lengths for all simulated cylinders while the error values of $AD(\hat{C})$ and $A(\theta)$ are very low.

3.2.2 Robustness in the Presence of Outliers: To see the robustness of the proposed approach, i.e., influence of the presence of outliers in cylinder fitting, we create 1000 quarter cylinder datasets with different percentages: 1%, 5%, 10%, 15%, 20% and 25%, of clustered outliers. We calculate performance measures as for the previous experiments for all the cases of outlier variations. Results are portrayed on the line diagrams in Fig. 4(a), (b), (c) and (d) for outlier percentages versus $AD(\hat{C})$, $A(\hat{R})$, $A(\hat{L})$ and $A(\theta)$, respectively. From the construction of the cylinder datasets and the definition of the performance measures it follows that the line diagrams for $AD(\hat{C})$, $A(\hat{R})$, $A(\hat{L})$ and $A(\theta)$ should be along virtual lines of 0, 1, 10 and 0, respectively. Fig. 4 shows that none of the existing methods follow the lines accordingly, when the outlier percentages increase. However, RLTS produces lines that are most close to 0, 1, 10 and 0 lines, almost independent of the percentage of outliers. That means that the proposed method is indeed not influenced by outliers.
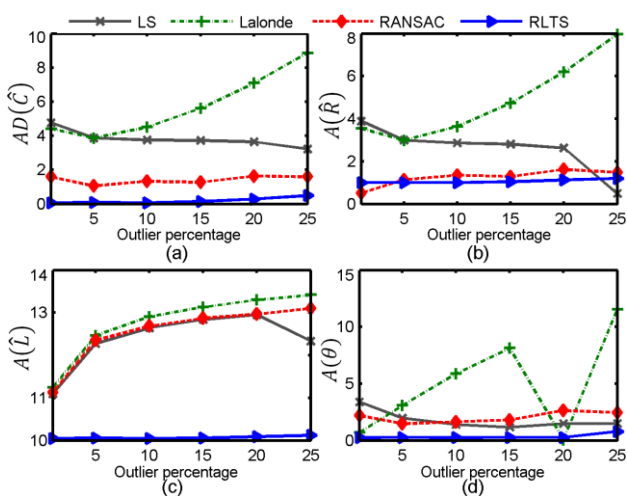

Figure 4. Influence of different percentages (1%,5%,10%,15%, 20% and 25%) of clustered outliers on cylinder fitting methods; line diagrams for outlier percentages versus performance measures (a) $AD(\hat{C})$, (b) $A(\hat{R})$, (c) $A(\hat{L})$, and (d) $A(\theta)$

In Fig. 5, box plots of $\hat{C}$, $\hat{R}$, $\hat{L}$ and θ are drawn for different methods to evaluate the robustness of the estimates based on 1000 values from 1000 samples with 10% clustered outliers. Boxes for the existing methods are larger than RLTS, which means their estimates are not as regular as RLTS. Moreover, existing methods have many outlying estimates (red '+' signs over and/or below the boxes). RANSAC has the maximum number of outlying results, whereas the new algorithm produces most reliable and robust results with very few outlying cases.
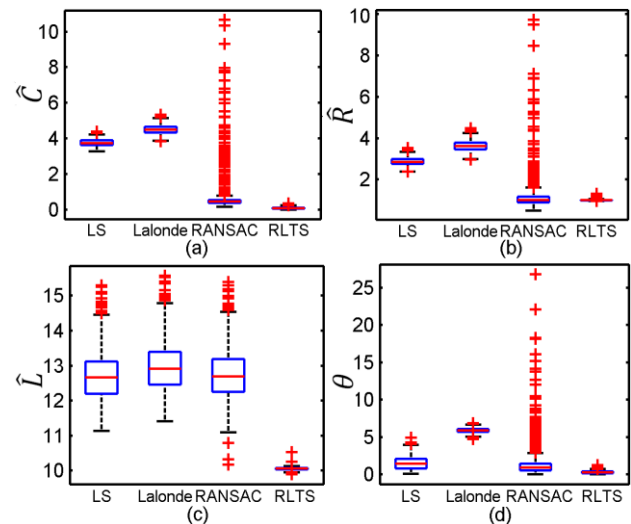

Figure 5. Robustness of the estimates from different cylinder fitting methods; Box plots for performance measures (a) $\hat{C}$, (b) $\hat{R}$, (c) $\hat{L}$, and (d) θ

3.2.3 Consistency of the Estimates: An estimator is said to be statistically consistent when with increasing sample size the estimates are closer to the real parameters. To check the consistency, we generate quarter cylinder datasets with increasing number of points n =100, 1000 and 10000 including 10% outliers. We fit cylinders using all methods and calculate the performance measures accordingly, results are in Table 2. For the existing methods, we do not see acceptable improvements. Even in some cases results are worse when the sample size is larger, for example LS, Lalonde and RANSAC fit cylinders with average lengths $A(\hat{L})$ =11.16m, 11.42m and 11.22m for n =100, and $A(\hat{L})$ = 13.87m, 14.14m and 13.92m for n =10000, respectively. For the proposed method all estimates $AD(\hat{C})$, $A(\hat{R})$, $A(\hat{L})$, and $A(\theta)$ are more accurate with increasing values of n =100, 1000 and 10000, while the error value $A(\theta)$ = 0.82, 0.24 and 0.09 is decreasing with increasing n =100, 1000 and 10000 respectively.

3.2.4 Cylinders with Different Radii: We simulate cylinders of different radii. Although we change the length of the radius, we simulate cylinders with a fixed number (100) of points. Therefore, the variations among the cylinders' radii can also be considered as the point density variations, when the point density is the number of points in a specific area that depends on the radius of a cylinder. To examine the effect of point density variations and the different sizes of radius on the estimates from different methods, we create datasets of different radii i.e., of different densities. 1000 quarter cylinder datasets including 10% clustered outliers are simulated for each of the radii 0.05m, 0.10m and 0.50m and with cylinder length 10m. The considered algorithms are executed on all datasets, and based on the results, the performance measures $AD(\hat{C})$, $A(\hat{R})$, $A(\hat{L})$ and $A(\theta)$ are calculated for all sizes of radii. Table 3 contains the results. The table shows that LS produces the worst results for $R = 0.05$ in terms of $AD(\hat{C})$, $A(\hat{R})$ and $A(\theta)$, while Lalonde produces the worst results for $R = 0.10$ and 0.50 in terms of $AD(\hat{C})$, $A(\hat{R})$ and $A(\hat{L})$. RANSAC as a robust approach performs remarkably better than LS and Lalonde, but still is not acceptable for reliable cylinder fitting. Results show that LS, Lalonde and RANSAC do better with a radius of 50cm than other sizes, in other words they are not good for smaller radius. For example, RANSAC produces $AD(\hat{C})$=119.21 and $AD(\hat{C})$=1.19 for R=5cm and 50cm, respectively. The proposed algorithm produces almost the same results with the change of radius. That means, RLTS have insignificant influence on changes of radius and point density variations.

3.3 Real Point Cloud Datasets

This section evaluates the proposed algorithm, and compares it with existing methods through assessment of two real point cloud datasets.

3.3.1 Signpost Data: Fig. 6(a) shows a road side signpost point cloud, collected by a survey company in Australia using vehicle borne MLS system at typical traffic speed. We call it signpost data, which is the same data as shown in Fig. 1. To fit a cylinder, we slice a rectangular portion in Fig. 6(a), shown in Fig. 6(b), which contains 231 points. A top view of Fig. 6(b) is shown in Fig. 6 (c). We see that approximately half of the cylinder is scanned. Some points indicated by the red ellipses in Fig. 6 (c) are clearly outliers, and do not match with the

majority of the circular points. We fit cylinders using LS, Lalonde, RANSAC and RLTS in Fig. 6 (e–h) respectively. Calculated radii and lengths are in Table 4. According to the PCA based approach, we get 2D projected circular points in Fig. 6(d), and the circles for Lalonde and RLTS as they use PCA. Table 4 and Fig. 6(d) show that Lalonde makes a larger circle with radius 10.1cm that considers many outlying points. On the other hand RLTS draws a circle with radius 5.8cm that follows the majority of the points, and successfully avoids outliers. It is closer to circular arc and match with the reality. Fig. 6 (g) shows that RANSAC does comparatively better than LS and Lalonde, still it fits a larger cylinder radius and length of 6.9cm and 2.884m respectively. Fig. 6 (d,h) explore that fits from the proposed algorithm are superior to any other existing method as most of the outlying points are ignored, and the outliers stay outside the fitted cylinder, while for the existing methods most of the outliers are accommodated by the fitted cylinders in Figs. 6(e–g). Fig. 6(i), which shows all fitted cylinders in one figure, clearly expresses that existing methods overestimate radius and length, mainly because the existing methods are negatively affected by outliers, or were unable to deal properly with incomplete sampling of the cylinder boundary.

3.3.2 Bridge-Pole Data: For this experiment we took a TLS scan sampling a high-way passing over a bridge, and picked the black part (a pole) in Fig. 7(a) to further demonstrate the cylinder fitting algorithms. Fig. 7(b), the black part of the bridge, which is cylindrical, has 4,282 points and includes some structures (part of cables) around the pole considered as outliers. We name the data bridge-pole. We perform all the considered cylinder fitting methods. To get the cylinder radius, as we are using a PCA based approach, we project the data onto a plane spanned by two PCs, and fit circles to the resulting 2D data. Results for Lalonde and RLTS are in Fig. 7(c). The circle obtained by Lalonde is much bigger than the RLTS circle. For better understanding, we extract only the RLTS circle in Fig. 7(d), which shows that RPCA based RLTS fits accurately to the 2D circular points and rightly avoids the points from the surrounding outlying structures. On the other hand, PCA based Lalonde fits a circle that is influenced by the outliers and therefore overestimates the cylinder parameters. We also fit cylinders using the LS and RANSAC methods, resulting fitted cylinders are shown in Figs. 7(e) and (g) respectively. LS fits a cylinder with a comparatively smaller radius (15.32m) than

Methods	n=100				n=1,000				n=10,000			
	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$
LS	3.88	2.98	11.16	2.60	3.77	2.88	12.67	1.52	3.74	2.85	13.87	1.37
Lalonde	4.66	3.77	11.42	5.90	4.52	3.65	12.93	5.89	4.49	3.61	14.14	5.84
RANSAC	0.85	1.27	11.22	1.63	0.73	1.21	12.71	1.37	0.59	1.14	13.92	1.26
RLTS	0.37	1.11	9.80	0.82	0.09	1.01	10.06	0.24	0.03	1.00	10.02	0.09

Table 2. Evaluation of estimated parameters for cylinders of different number of points
[$AD(\hat{C})$ = Average Distance between the real cylinder centres and the fitted cylinder centres, $A(\hat{R})$ = estimated Average radius, $A(\hat{L})$ = estimated Average length, and $A(\theta)$ = Average bias orientation]

Methods	R = 0.05				R = 0.10				R = 0.50			
	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$	$AD(\hat{C})$	$A(\hat{R})$	$A(\hat{L})$	$A(\theta)$
LS	624.16	553.22	12.16	14.78	592.08	529.96	11.96	16.01	3.18	2.39	12.16	9.36
Lalonde	390.46	390.45	13.11	7.19	630.54	630.52	13.09	7.08	7.68	7.28	13.00	6.46
RANSAC	119.21	2.20	12.83	2.19	6.06	3.71	12.76	2.92	1.19	1.01	12.75	1.55
RLTS	0.09	0.06	10.05	0.04	0.14	0.09	10.08	0.04	0.10	0.51	10.06	0.14

Table 3. Evaluation of estimated parameters for cylinders with different radii (in metre)
[$AD(\hat{C})$ = Average Distance between the real cylinder centres and the fitted cylinder centres, $A(\hat{R})$ = estimated Average radius, $A(\hat{L})$ = estimated Average length, and $A(\theta)$ = Average bias orientation]
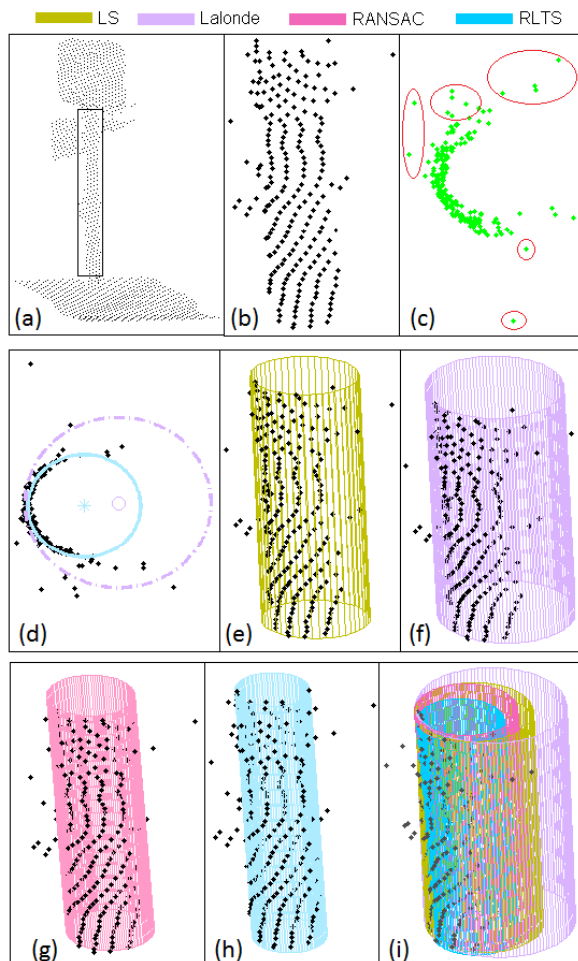
Figure 6. Cylinder fitting for signpost data; (a) signpost data, (b) 3D PCD for cylinder fitting, (c) top view of the cylinder data, (d) fitted circles for Lalonde and RLTS, (e) fitted cylinder for LS with real points (black), (f) fitted cylinder for Lalonde, (g) fitted cylinder for RANSAC, (h) fitted cylinder for RLTS, and (i) all the fitted cylinders from different methods in a same figure

Methods	Radius	Length
LS	0.079	2.885
Lalonde	0.101	2.885
RANSAC	0.069	2.884
RLTS	0.058	2.865

Table 4. Estimated radii (m) and lengths (m) of the fitted cylinders from different methods for signpost data

Lalonde (18.72m) but with a remarkably larger radius than RANSAC (0.49m) and RLTS (0.45m). If we consider the estimated lengths for the fitted cylinders, results in Table 5 show that all the existing methods overestimate the cylinder's length. Fig.7 (g) indicates that RANSAC is influenced by outliers as the fitted cylinder (red) is masked from sight from the real points of the data (black). It also shows that the length of the cylinder is too high as more of the cable (steel wire) points are included in the data. That means, RANSAC is influenced by outliers, as it considers outliers as inliers which is the well-known masking effect. In contrast, the proposed approach fits cylinder in Fig. 7(h) more accurately than any other existing methods resulting in a length of 12.65m, as the fitted cylinder fits all the inliers properly.
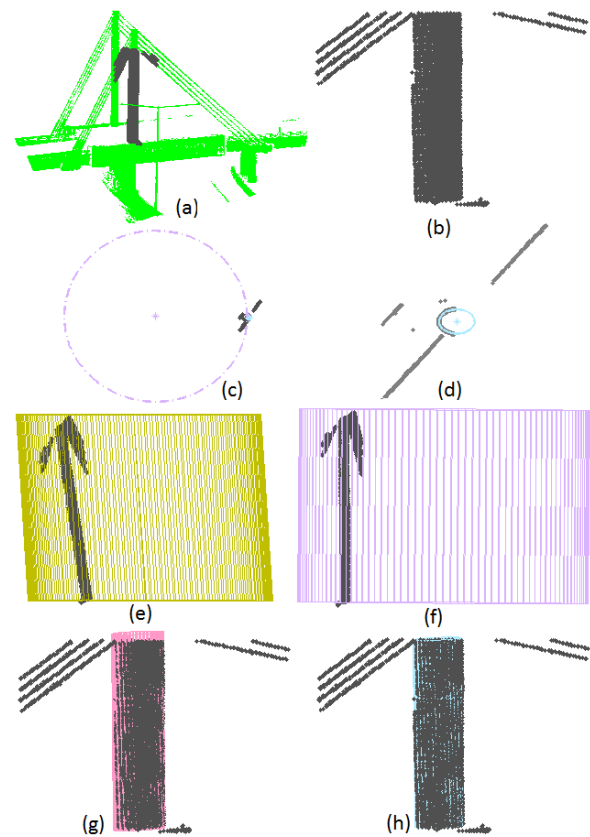

Figure 7. Cylinder fitting for bridge-pole data; (a) bridge-pole data in a scene, (b) bridge-pole for cylinder fitting, (c) fitted circles for Lalonde (bigger one) and RLTS onto the 2D RPCA based plane, (d) 2D points and fitted circle by RLTS, (e) fitted cylinder (olive) for the LS with real points (black), (f) fitted cylinder (violet) for Lalonde, (g) fitted cylinder for RANSAC (pink), and (h) fitted cylinder (feroza) for RLTS

Methods	Radius	Length
LS	15.32	13.39
Lalonde	18.72	13.32
RANSAC	0.49	13.27
RLTS	0.45	12.65

Table 5. Estimated radii (m) and lengths (m) of the fitted cylinders from different methods for bridge-pole data

4. CONCLUSIONS

In this paper, a robust statistical approach for cylinder fitting is proposed for 3D PCD. The new method successfully fits cylinders, in case of cylinder data contaminated by outliers, for only partially scanned cylinders as well as for full data. Several experiments using artificial and real laser scanning data exhibit that the new approach is robust towards a high percentage of clustered outliers and noise. It is statistically consistent and robust when the size of the cylinder's radius and point density is varying. It produces a very small amount of errors when the position of the cylinder and orientation is changed. For example, for 1000 simulated quarter cylinders with 10% clustered outliers, Lalonde (a PCA based approach) fits cylinders with on average orientation error θ of 5.87º, whereas the new RPCA based algorithm fits cylinder with on average orientation error of only 0.36º. In terms of radius Lalonde and RLTS fit cylinders on average 3.63 and 1.02 times bigger, respectively, than the real radius of 1m. Among the PCA based methods, the method of Lalonde over fits because it is using non-robust PCA and therefore partial incorrect PCs. The new algorithm has potentials

for many applications such as in street furniture management, forest inventory, structural health monitoring and machine tools quality control. Further study will consider data with more inlier variation, cylinders with different diameters at their ends, and large-scale data sampling complex scenes exhibiting multiple cylinders.

ACKNOWLEDGEMENTS

REFERENCES

Ahmed, M., Haas, C., and Haas, R., 2014. Automatic detection of cylindrical objects in built facilities. *J. Comput. Civ. Eng.*, 28(3), pp. 1943–5487.

Al-Sharadqah, A., and Chernov, N., 2009. Error analysis for circle fitting algorithms. *Electron. J. Stat.,* 3, pp. 886–911.

Beder, C., and Förstner, W., 2006. Direct solutions for computing cylinders from minimal sets of 3d points. 9th *European Conference on Computer Vision,* 7–13 May, Graz, Austria, pp. 135–146.

Chaperon, T., and Goulette, F., 2001, Extracting cylinders in full 3D data using a random sampling method and the Gaussian image. *Proc. of the Vision Modeling and Visualization Conference*, 21–23 November, Stuttgart, Germany, pp.35–42.

Deschaud, J. -E., and Goulette, F., 2010. A fast and accurate plane detection algorithm for large noisy point clouds using filtered normals and voxel growing. *Proc. Intl. Symp. 3DPVT*, Paris.

De Guevara, I. L., Muñoz, J., de Cózar, O. D., Blázquez, E. B., 2011. Robust fitting of circle arcs. *J. Math. Imaging and Vision*, 40(2), pp. 147–161.

Duda, R.O., and Hart, P.E., 1972. Use of Hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15 (1), pp. 11–15.

Fischler, M. A., and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM*, 24(6), pp. 381–395.

Franaszek, M., 2012. Variances of cylinder parameters fitted to range data. *Journal of Research of the National Institute of Standards and Technology*, 117, pp. 257–267.

Gosliga van R., Lindenbergh, R., and Pfeifer, N., 2006. Deformation analysis of a bored tunnel by means of terrestrial laser scanning. IASPRS Vol. XXXVI/5, Dresden 25-27 September.

Hubert, M., Rousseeuw, P. J., and Branden, K. V., 2005. ROBPCA: a new approach to robust principal component analysis, *Technometrics*, 47(1), pp. 64–79.

Kwon, S. W., Bosche, F., Kim, C., Haas, C. T., Liapi, K. A., 2004. Fitting range data to primitives for rapid local 3D modeling using sparse range point clouds. *Automation in Construction*, 13(1), pp. 67–81.

Lalonde, J. F., Vandapel, N., and Hebert, M., 2006. Automatic three–dimensional point cloud processing for forest inventory. CMU-RI-TR-06-21,The Robot. Inst. CMU, Pennsylvania, USA.

Marshall, D., Lukacs, G., and Martin, R. (2001). Robust segmentation of primitives from range data in presence of geometric degeneracy. *IEEE Trans. PAMI.*, 23(3), pp. 304–314.

Méndez, V., Rosell-Polo, J. R., Sanz, R., Escolà, A., Catalán, H., 2014. Deciduous tree reconstruction algorithm based on cylinder fitting from mobile terrestrial laser scanned point clouds. *Biosystems Engineering*, 124, pp. 78–88.

Nurunnabi, A., Belton, D., and West, G., 2014. Robust statistical approaches for local planar surface fitting in 3D laser scanning data. *ISPRS J. Photogramm. Remote Sens.*, 96, pp. 106–122.

Nurunnabi, A., West, G., and Belton, D., 2015. Outlier detection and robust normal-curvature estimation in mobile laser scanning 3D point cloud data. *Pattern Recognit.* 48(4), pp. 1404–1419.

Nurunnabi, A., Belton, D., West, G., 2016. Robust segmentation for large volumes of laser scanning 3D point cloud data. *IEEE Trans. Geosci.and Remote Sens.*54(8), pp. 4790–4805.

Rabbani, T., and Van Den Heuvel, F., 2005. Efficient Hough transform for automatic detection of cylinders in point clouds. ISPRS WG III/3,4, V/3 Workshop "Laser Scanning", pp. 60–65.

Rahayem, M., Werghi, N., Kjellander, J., 2012. Best ellipse and cylinder parameters estimation from laser profile scan sections. *Optics and Lasers in Engineering*, 50(9), pp. 1242–1259.

Rousseeuw, P. J., and van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), pp. 212–223.

Rousseeuw, P.J., 1984. Least median of squares regression. *J. Amer. Statist. Assoc.*, 79(388), pp. 871–880.

Rousseeuw, P.J., and Leroy, A., 2003. *Robust Regression and Outlier Detection*, John Wiley and Sons, New York, USA.

Schnabel, R., Wahl, R., and Klein, R., 2007. Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum*, 26(2), pp. 214–226.

Shakarji, M. C., 1998. Least squares fitting algorithms of the NIST algorithm system. *Journal of Research of the National Institute of Standards and Technology*, 103(6), pp. 633–641.

Su, Y-T., Bethel, J., and Hu, S., 2016.Octree-based segmentation for terrestrial LiDAR point cloud data in industrial applications. *ISPRS J. Photogram. Remote Sens.,* 113, pp. 59–74.

Tarsha-Kurdi, F., Landes, T., and Grussenmeyer, P., 2007. Hough-transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from LiDAR data. IAPRS, Vol. XXXVI/3, pp. 407–412.

Tran., T-T., Cao, V-T., and Laurendeau, D., 2015. Extraction of cylinders and estimation of their parameters from point clouds. *Computers & Graphics*, 46, pp. 345–357.

Vosselman, G., Gorte, B. G. H., Sithole, G., and Rabbani, T., 2004. Recognizing structure in laser scanner point clouds. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(Part 8/W2): pp. 33–38.

Wang, D., Hollaus, M., Puttonen, E., Pfeifer, N., 2016. Fast and robust stem reconstruction in complex environments using terrestrial laser scanning data. IAPRS,Vol. XLI-B3, pp.411–417.

Wang, H., and Suter, D., 2003. Using symmetry in robust model fitting. *Pattern Recognit. Lett.*, 24(16), pp. 2953–2966.