Robust cylinder fitting in laser scanning point cloud data

Abdul Nurunnabi, Yukio Sadahiro, Roderik Lindenbergh, David Belton

PII: S0263-2241(19)30104-6
DOI: https://doi.org/10.1016/j.measurement.2019.01.095
Reference: MEASUR 6339

To appear in: *Measurement*

Received Date: 5 December 2018
Revised Date: 28 January 2019
Accepted Date: 31 January 2019

Please cite this article as: A. Nurunnabi, Y. Sadahiro, R. Lindenbergh, D. Belton, Robust cylinder fitting in laser scanning point cloud data, *Measurement* (2019), doi: https://doi.org/10.1016/j.measurement.2019.01.095

Robust cylinder fitting in laser scanning point cloud data

Abdul Nurunnabi [a,*], Yukio Sadahiro [b], Roderik Lindenbergh [c], and David Belton [d]

[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
[d]Department of Spatial Sciences, Curtin University, Perth, Western Australia, Australia
[a,b]{a.nurunnabi,sada}@csis.u-tokyo.ac.jp, [c]r.c.lindenbergh@tudelft.nl, [d]d.belton@curtin.edu.au

ABSTRACT

Cylinders play a vital role in representing geometry of environmental and man-made structures. Most existing cylinder fitting methods perform well for outlier free data sampling a full cylinder, but are not reliable in the presence of outliers or incomplete data. Point Cloud Data (PCD) are typically outlier contaminated and incomplete. This paper presents two robust cylinder fitting algorithms for PCD that use robust Principal Component Analysis (PCA) and robust regression. Experiments with simulated and real data show that the new methods are efficient (i) in the presence of outliers, (ii) for partially and fully sampled cylinders, (iii) for small and large numbers of points, (iv) for various sizes: radii and lengths, and (v) for cylinders with unequal radii at their ends. A simulation study consisting of 1000 cylinders of 1m radius with 20% clustered outliers, reveals that a PCA based method fits cylinders with an average radius of 2.84m and with a principal axis biased by outliers of 9.65^0on average, whereas the proposed robust method correctly estimates the average radius of 1m with only 0.27^0 bias angle in the principal axis.

Keywords: *3D Modelling, Feature Extraction, Robust Measurement, Robust PCA, Robust Regression, Shape Reconstruction*

1. Introduction

Cylinder fitting is important for many applications of 3D point clouds, and can also be incorporated in several (pre)processing methods. Examples of applications and methods include object recognition and modelling [1], feature extraction [2], surface reconstruction [3], registration [4], industrial applications [5], engineering and planning [6, 7], autonomous navigation [8], archaeological documentation [9], orthopaedics [10], manufacturing and quality control [11, 12], as-built modelling [13], instrument calibration [14], and forest inventory [15, 16]. Therefore, cylinder fitting directly relates to relevant subjects such as computational metrology, augmented reality, computer vision, computer graphics, nuclear physics, reverse engineering, pattern recognition, photogrammetry and remote sensing. A successful and robust cylinder fitting approach could be integrated within a laser based street inventory workflow to monitor the geometric state of, for example, light poles and street signs [17]. Airborne, Terrestrial and Mobile

* Corresponding author.

Laser Scanning (i.e., ALS, TLS and MLS) systems have been accepted as an efficient technology for high-speed direct acquisition of dense three-dimensional (3D) spatial measurements known as point clouds that can potential come with millimetre precision and completeness in surface sampling. Urban areas, road corridors, electric installations, industrial and archaeological sites and forests all feature cylindrical shapes e.g., as part of street light/utility/sign poles, pipes, bollards, sleeves, connectors and trees. Fitting geometric shapes in Point Cloud Data (PCD) is challenging because the data are usually incomplete, unordered, biased, do not follow any statistical distribution, have uneven point density and are locally sparsely populated. Moreover, outliers that do not follow the pattern of the majority of points are common in point clouds, mainly because of systematic biases and sensor errors, multipath reflections and occlusions. Moving objects that pass through the scan area faster than they can be captured like snow, rain, dust, and noise may appear as off-surface points and act as outliers [18]. In addition, multiple and unorganized structures in a data set may create clustered and pseudo-outliers [18, 19]. For example, points sampling a wall in front of a lamp post can act as pseudo outliers when the lamp post is under study. The presence of outliers aggravates the complexity of fitting shapes in PCD, and leads to a high measurement inconsistency.

Existing cylinder fitting methods have focused mainly on full (complete) data, i.e., it is assumed that cylindrical objects are fully scanned. As far as we know, no comprehensive study so far has been conducted on cylinder fitting in incomplete PCD contaminated by outliers. In this paper, we are motivated by point clouds obtained by laser scanner that often contain incomplete cylindrical surfaces. For example, mobile laser scanning often results in partly sampled cylindrical street poles. Indeed, parts of poles are not visible for the scan system as it typically operates only from the street side. In addition, in complex scenes, objects obstruct each other. In such cases, poles or tree trunks are only partly scanned; resulting in an incomplete sample of a cylindrical shape, demonstrated in Fig. 1. Except for the incompleteness, points sampling the edge of a cylindrical object w.r.t. the laser scanner position are expected to have lower quality, because of the almost tangential measurement geometry. Also in industrial settings and at archaeological sites it is often not possible to scan objects from different positions because of the scene complexity. So, robust cylinder fitting in incomplete PCD with outliers has great importance.

Most methods for geometric primitive (e.g., planes and cylinders) extraction and fitting are based on either Least Squares (LS) and/or Singular Value Decomposition (SVD), which are both sensitive to outliers and not statistically robust. The well-known RANdom SAmple Consensus (RANSAC; [20]) method has been used for robust model fitting in the presence of outliers. A pioneer work on using RANSAC for cylinder fitting in range data was presented by Bolles and Fischler in [21]. However, the method presented in [21] is for depth data or 2.5D data. At a later stage, surface fitting approaches were used to find cylinders in 3D data [22]. Lee et al. [23] developed a RANSAC based method for extracting cylindrical pipelines in point cloud data of a plant. We will see later that RANSAC is not sufficiently free from outlier effects. Faber and Fisher [24] addressed the problem of cylinder fitting to 3D data by a constrained Euclidean fitting method. The authors [24] claimed that their method produces better results compared to the more generally used algebraic fitting formulation. Hopkinson et al. [25] used LS regression for estimating stem diameters by circle fitting. We know LS regression is sensitive to outliers. Wang and Suter [26] successfully used robust regression (cf., [27]) coupled with the use of a symmetric distance defined in [28] for robust circle and ellipse fitting in 2D image data. The authors pointed out however that their method is not a good choice for spatially asymmetric (i.e., partial) data. Lalonde et al. [15] followed Kwon et al. [1]'s idea of using Principal Component Analysis (PCA) for cylinder fitting and applied it for tree trunk estimation. Circle-fit based methods including [15] have been used in many approaches for cylinder fitting. The authors [15] demonstrated that their PCA based approach is significantly faster and more accurate than other approaches, but, they did not consider the presence of outliers. It is widely recognized that PCA is highly sensitive to outliers and may produce misleading outputs in the case of point

cloud processing (cf., [19, 29]). Forsman et al., [30] found out that diameters of cylindrical tree stems are often overestimated for data collected by terrestrial laser scanning. The authors [30] developed a compensation function to reduce the bias that may arise when estimating cylinder diameters starting from a circle-fitting method.

To overcome these issues: sensitivity to outliers and incompletely sampled cylinders, this paper proposes and compares two robust approaches for cylinder fitting; both of which are an extension of an algorithm as introduced in [31]. The two new algorithms use Robust PCA (RPCA), robust regression and incorporate a circle fitting method recently proposed in [32]. As a bi-product the proposed methods do enhance the performance of the Hyper algorithm for robust circle fitting. The methods are efficient for partial (incomplete) cylinder data as well as for full (complete) cylinder data in the presence of noise and up to 25% of clustered and 70% of scattered outliers. Moreover, the new methods are statistically consistent and produce robust parameters in case of varying number of data points, different sizes (radii and lengths) of cylinders and varying data densities. The methods successfully find radii even if these are unequal at the cylinder ends. The main technical difficulty is that, as the new methods follow an iterative approach, they are, as may be expected, computationally more intensive compared to classical, non-iterative methods based on linear LS and PCA. In addition, the newly developed methods require additional pre-processing of the raw input PCD to identify data corresponding to a cylindrical shaped object.

The rest of this paper is arranged as follows. Related previous studies, technical background and principles used in the proposed algorithms, and the methods used for comparison are briefly discussed in Section 2. Two new algorithms are proposed in Section 3. Experiments on artificial and real point clouds are analysed and evaluated in Section 4. A brief discussion on why and how the proposed algorithms work is presented in Section 5, followed by conclusions in Section 6.

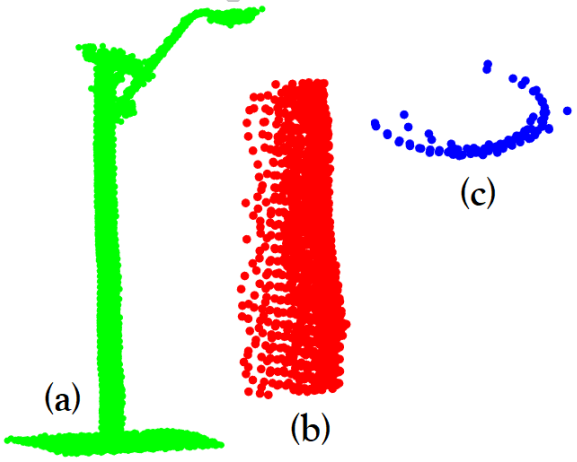

Fig. 1. (a) Mobile laser scanning point cloud data; a utility pole, (b) a cylindrical segment of the pole, and (c) a horizontal cross-section at the top of the cylinder end of the segmented pole consisting of an incomplete circular arc.

2. Literature review, related principles and methods used for comparison

This section consists of three subsections, one on existing cylinder fitting methods, one describing basic principles and methods used in the proposed algorithms, and the last consisting of short descriptions of the methods used in comparison for the new algorithms.

2.1. Existing methods

Various methods for fitting cylinders in 2/3D data have been developed over the years that we divide into two main categories: (i) methods working on raw data, which means that the fitting is performed before any type of relevant labelling process like classification or segmentation took place, and (ii) methods based on labelled (cylindrical) data.

2.1.1. Methods working on raw data

The methods, which working on raw data are commonly known as voting techniques, and are based on Hough Transformation (HT; [33]), RANSAC or clustering (cf., [10]). Hough Transformation and RANSAC have success stories, but many people are reluctant to use them because these methods are computationally expensive and allocate huge memory, notably when the number of parameters to be estimated is above two or three. Rahayem et al. [8] pointed out that clustering methods take much processing time, need massive memory and are relatively less accurate. Beder and Forstner [11] provided a direct solution for estimating circular cylinder parameters in range data, but one of the problems of their method is that the maximum number of solutions for random point configurations is unknown. Schnabel et al. [3] developed an efficient RANSAC algorithm for shape detection that includes cylindrical shape. The authors mentioned that their method can be applied to a large variety of data and is highly robust, but does not find shape proxies for every part of the surface. In investigation of plane extraction, Deschaud and Goulette [34] argued that RANSAC is efficient for detecting large planes in noisy point clouds, but very slow for extracting small planes in large point clouds. Chaperon and Goulette [35] proposed a RANSAC based two-step approach for cylinder fitting, in which the first step finds the cylinder's direction in a so-called Gaussian image while the second step finds the cylinder's position and size. In each step they used random sampling of minimal subsets. Rabbani and Heuvel [6] pointed out that this method is limited in the presence of multiple cylinders of different radii but with similar orientation. Vosselman et al. [2] introduced a HT based cylinder detection algorithm that uses point normals to estimate the cylinder orientation. Rabbani and Heuvel [6] developed a two-step sequential HT for automatic detection of cylinders in point clouds. Some studies (cf., [34]) showed that HT is less efficient than RANSAC in terms of computation time when a model is fitted in large data, and is sensitive to the segmentation parameters when using segment based point cloud processing.

2.1.2. Methods starting from labelled cylinder data

The second category of methods starts by segmenting sampled objects as much as possible so that every segment directly corresponds to an appropriate geometric shape. This category often needs manual supervision during modelling, and its success may depend on the success of pre-processing (e.g., segmentation). This type of methods fits cylinders to labelled (segmented or classified) data i.e., it fits a cylinder to data likely to be cylindrical. Existing methods can be further grouped as non-linear and linear LS methods [8, 36]. Linear LS methods are based on a linearization of the Euclidean distance between the points and the modelled surface. LS based geometric fitting is simple, generic and has many favourable mathematical properties. It is however sensitive to outliers [37] as it is typically assumed that errors are normally distributed. Some authors (cf., [38]) suggested using a Gaussian mixture model to mitigate the effect of outliers to the errors distribution. Liang et al. [16] used a robust weighting function in their cylinder fitting algorithm to reduce the effect of outliers when minimizing the sum of squared error residuals, which is the objective function when the LS principle is used. But the authors did not consider cases where cylinders were only partially sampled. Palancz et al. [39] developed an extended form of vector algebraic modelling for cylinder parameters estimation. Kawashima et al. [40] used Least Median Squares (LMS) principle together with the Levenberg-Marquardt method [41] to fit cylinders for recognizing piping systems. Non-linear methods, also known as geometric fitting, minimize the orthogonal distances of points to the surface of the

cylinder (cf., [42]). Franaszek [43] studied the variances of cylinder parameters for non-linear LS methods. Al-Subaihi [44] proposed a cylinder fitting method based on the minimization of the sum of orthogonal squared distances, but the method is limited to full circle data without outliers.

Kwon et al. [1] used PCA for cylinder fitting without using the normals of the data, and they claimed their method has reasonable accuracy for modelling objects at construction sites. Lalonde et al. [15] employed a PCA based approach and developed two algorithms based on 2D projection and 3D fitting for tree trunk segmentation. The authors [15] showed that results from the 3D fitting technique were approximately 25% worse than the PCA based 2D projection method. The authors in both [1, 15] claimed that PCA based methods are computationally efficient, but they did not consider data with gross or pseudo-outliers. It is evident that PCA is highly sensitive to outliers in point cloud processing (cf., [19, 29]). Faber and Fisher [24] studied various distance functions and proposed improved versions of cylinder fitting by using closed form expressions of the real Euclidean distance. The main disadvantage of Euclidean fitting is the associated computational cost, though the authors [24] remarked that computational efficiency has become less important due to increasing computing speed. Rahayem et al. [8] used ellipse fitting prior to cylinder fitting, but did not consider the presence of outliers. Some studies (e.g., [12, 23, 40]) concentrate on cylinder detection and are able to extract multiple cylinders successfully in complex scenes using conventional LS, RANSAC or algebraic techniques.

The starting point of this paper is that classical methods are hampered by outlier effects, which may produce significant errors, notably when used in automated work-flows without supervision. The reader is referred to [3, 6, 12, 22, 39, 41, 42, 43] for more information on cylinder fitting.

2.2. Related principles, methods and issues

The algorithms developed in this paper use robust statistical approaches: robust PCA and robust regression to get statistically robust estimates of cylinder parameters. This section outlines the basic premise of PCA, robust version of PCA, LS and robust regression.

2.2.1. PCA and robust PCA

Principal Component Analysis (PCA) is a statistical technique that is able to find directions of data variability by determining orthonormal axes, which maximally de-correlate the dataset (cf., [45]). It has been used in many ways for point cloud processing [12, 46]. To get Principal Components (PCs), Singular Value Decomposition (SVD) is performed on the covariance matrix Σ, which is defined as:

$$\Sigma_{3\times 3} = \frac{1}{k}\sum_{i=1}^{k}(p_i - \overline{p})(p_i - \overline{p})^T, \quad (1)$$

where $p_i = (x_i, y_i, z_i)^T; i = 1,2,...,k$ and $\overline{p} = (\overline{x}_i, \overline{y}_i, \overline{z}_i)^T$ are the i-th point and the centre (mean) of PCD consisting of k points, respectively. PCA generates two matrices; one is a diagonal matrix with the three eigenvalues $(\lambda_i; i = 2,1,0; \lambda_2 \geq \lambda_1 \geq \lambda_0)$ of Σ as its diagonal elements, while the other contains corresponding eigenvectors (i.e., PCs; v_2, v_1 *and* v_0). PCs rank data variability in terms of eigenvalues, and produce the corresponding directions. The first PC (PC1) v_2 directs the highest variability of the data, and so forth: PC2 (v_1) and PC3 (v_0). It is known that PCA is extremely sensitive to outliers (cf., [47]). Nurunnabi et al. [29] used robust versions of PCA (RPCA) in point cloud analysis, and showed that the RPCA introduced by Hubert et al. [48] is a successful robust alternative to the classical PCA that minimizes the effects of outliers and produces reliable PCs. The authors [48] followed the principle of finding multivariate outliers: if a point is an outlier then there must be some one-dimensional projection of the data for which the point is also a univariate outlier. Hence, an 'outlyingness' measure is derived to identify outliers by projecting the data

values on many univariate directions. In each direction every point is scored by its corresponding value of being an outlier, where the outlyingness of the *i*-th point in the *v* direction is calculated as

$$w_i = argmax_v \frac{|p_i v^T - c_{\mathrm{FMCD}}(p_i v^T)|}{\Sigma_{\mathrm{FMCD}}(p_i v^T)}, \quad (2)$$

where $p_i v^T$ denotes the projection of the *i*-th point onto the *v* direction, and c_{FMCD} and Σ_{FMCD} are the Fast Minimum Covariance Determinant (FMCD; [49]) based robust mean and covariance matrix in *v* direction. A majority portion *h* ($h > k/2$) of total *k* observations with the smallest outlyingness (w_i) values is used to build a robust covariance matrix Σ_h. Then the observations are projected onto the *d* dimensional subspace spanned by the *d* largest eigenvectors of Σ_h, and the mean vector and the covariance matrix are computed by means of the reweighted FMCD estimator, with weights based on the robust distance of every point. The eigenvectors of this covariance matrix built from reweighted observations are the required robust PCs. The authors [48] claimed that RPCA produces accurate estimates for data without outliers and more robust estimates for data with outliers, moreover, the resultant robust PCs are location and orthogonal invariant. The reader is referred to Hubert et al. [48] for more information on RPCA.

2.2.2. Least square and robust regression

Given *k* observations, the standard linear regression model is

$$Y = X\beta + \varepsilon, \quad (3)$$

where *X* is a $k \times (m+1)$ full rank design matrix of m explanatory variables, β is an $(m+1) \times 1$ vector of parameters, ε is a $k \times 1$ vector of random errors following a Gaussian normal distribution with mean 0 and variance σ^2, and Y is a $k \times 1$ vector of response variable. The LS approach estimates model parameters by minimizing the sum of the squared residuals, i.e.,

$$minimze_{\hat{\beta}} \sum_{i=1}^{k} e_i^2, \quad (4)$$

where the *i*-th residual is given by $e_i = Y_i - \hat{Y}_i$, $\hat{Y}_i = X_i^T \hat{\beta}$, and the estimated parameter vector is obtained by the expression $\hat{\beta} = (X^T X)^{-1} X^T Y$. LS is one of the most commonly used model parameter estimation methods, but it has been criticized for its sensitivity to outliers. Indeed, its Breakdown Point (BP; a robustness measure that reports the percentage of outliers that may cause unacceptable estimates), is zero [27]. To reduce outlier influence, Huber (cf., [50]) introduced so-called M-estimators. His idea is to replace the squared residuals in LS estimation by some function $\rho(e_i)$ of the residuals, i.e.,

$$minimze_{\hat{\beta}} \sum_{i=1}^{k} \rho(e_i), \quad (5)$$

where ρ is a positive-definite and symmetric function that typically has a unique minimum at zero. Although M-estimators are statistically robust, the BP of M-estimators is still zero in case of leverage points in *X*, i.e., points for which the x-coordinate is far from the bulk of the *X* in the data. Many other robust estimators like L-estimators, R-estimators and S-estimators have been introduced (cf., [51]). Rousseeuw [52] introduced the two most popular high-breakdown robust regression methods; the first one is Least Median Squares (LMS) that minimizes the median of squared residuals:

$$minimize_{\hat{\beta}}\ median_i e_i^2. \quad (6)$$

It has been used in many applications including conic fitting [53] and cylinder fitting [40]. The second approach is Least Trimmed Squares (LTS), which trims the $k - h$ observations having the largest residuals and consecutively minimizes the sum of the remaining *h* ($h > k/2$) least squared residuals as per Eq. (7)

$$minimize_{\hat{\beta}} \sum_{i=1}^{h} e_i^2. \quad (7)$$

The author [52] claimed that both LMS and LTS achieve 50% BP, when h is almost $k/2$. Hubert et al. [54] pointed out that LMS has a slow convergence rate and its asymptotic efficiency is zero, while inversely, LTS is asymptotically normal and can be computed faster. Another reason why LTS is preferred to LMS, is because LTS is locally stable while LMS is not, which means that in case of LMS an infinitesimal change in the data can greatly affect the output. Wang and Suter [26] used LTS for the robust fitting of circles and ellipses. Another popular approach is Weighted LS (WLS) which follows a diagnostic principle by, first, finding outliers and then, second, assigning a weight to each point according to the degree of that point being an outlier (outlyingness). In its optimization step, WLS minimizes the sum of the weighted least squared residuals. The reader is referred to [27, 51, 54] for more on robust and diagnostic regression.

2.3. Methods used for comparison

The new algorithms in this paper are compared to six existing methods. Two classic methods: LS and the PCA method of Lalonde et al. [15], two robust methods: RANSAC [20] and MSAC [55], and two recently proposed techniques: developed in Kawashima et al. [40], and Tran et al. [12]. In addition, the results of the new methods are compared to the results of the Cyclone [56] software from Leica for the real laser scanning data experiments.

The LS and Lalonde et al. [15] methods are not robust, but we consider them because the LS principle is used as a basis of many other methods (cf., [57]), while Lalonde et al. [15] has been recognized as one of the faster approaches using the PCA paradigm; it outperforms several existing methods. RANSAC is included to the comparison because it is the most popular robust technique (cf., [58]). It possesses excellent properties for object shape and geometric primitive detection, modelling, fitting and reconstruction, especially in case of cluttered scenes and/or in the presence of high percentages of outliers. RANSAC is an iterative process that randomly selects a minimal subset of points and calculates the parameters of the considered geometric primitive based on this subset, and then repeats the process many times. The score of one selection is the number of points in the full data set that supports this subset. Points supporting a subset are deemed inliers, and being an inlier is evaluated by comparing the distance of the point to an inlier-noise threshold. RANSAC selects the subset supported by the most inliers as the most probable. The final estimation is based on an LS adjustment using all these inliers. Torr and Zisserman [55] showed that RANSAC can be sensitive to the choice of the correct noise threshold and may be affected by the presence of multiple structures. Many improvements of RANSAC have been proposed (cf., [3, 55]). Choi et al. [59] evaluated a number of methods and showed that the M-estimator SAmple Consensus (MSAC; [55]) is one of the most accurate versions. It uses the well-known statistically robust M-estimator [50] as a bounded loss function, and inliers are scored according to their goodness-of-fit to the data. The authors [55] claimed that MSAC yields a modest to hefty benefit for all robust estimations with only slightly more computational effort. MSAC is commonly applied in point cloud processing (cf., [29, 60]).

Kawashima et al. [40] developed a cylinder fitting algorithm, which collects the normal vectors of the potential cylinder points on a unit Gaussian sphere. A RANSAC algorithm is employed for fitting a plane to the Gaussian points using a fixed number (50) of runs, while as outlier threshold; the inner product of normals is used at a value of 0.01. The normal vector to the fitted plane is used as an initial estimate of the cylinder axis. Next, LMS regression is used with a specific number (10) of iterations for fitting a circle to the cylinder points projected onto the plane. The estimated centre and radius are initial estimates for the axis and radius of the cylinder. Next, a cylinder is finally fitted to the points using the Levenberg-Marquardt (cf., [41]) algorithm. The reliability of the fit and the influence of the aberrant values to the parameters are checked using an angular threshold for a specific number of regions on the cylinder surface. The reader is referred to Kawashima et al. [40] for details.

Tran et al. [12] developed an iterative process for cylinder fitting that uses PCA based point normals (cf., [46]) as input for a covariance matrix. Performing PCA again on that covariance matrix estimates the PC3 (corresponding to the smallest eigenvalue) that is defined as the cylinder axis and allows to project the inlier points (neighbourhood of each potential cylinder point) onto the other two PCs based plane. A circle is fitted to the projected points using Pratt's method (an algebraic approach; [61]), and the estimated parameters: centre and radius of this circle are used as the required cylinder's centre and radius. Inliers are updated at each iteration (usually 10 times) in the process. The results from the 10^{th} iteration are used as the final fitting. The reader is referred to [12] for more about the algorithm.

Hereafter, for convenience, we refer to the methods by Lalonde et al. [15], Kawashima et al. [40] and Tran et al. [12] by the acronyms LVH, KKD and TCL, respectively, which are created from the first letters of the family names of the contributing authors for each paper.

3. Proposed algorithms and evaluation methodology

3.1. Problem formulation and new algorithms development

This section proposes two robust cylinder fitting algorithms. Assumed input for the algorithms is a point cloud sampling a cylinder. This means that in general a pre-processing step is required to extract cylindrical data from a large point cloud. Extraction could be done manually, by a human operator, that selects a cylindrical object using a bounding box, or by an automated workflow. For mobile mapping such workflows have been developed to extract road-side trees [62], or street furniture like lamp poles or traffic signs [17]. This type of sequential tasks, usually first identifies terrain points, cluster off –terrain points, decomposes clusters using e.g., spatial segmentation and identifies cylindrical parts in clusters containing one tree or traffic sign, cylindrical parts can be recognized as linear features or identified using a suitable segmentation method [6]. As an alternative to spatial segmentation region growing based robust segmentation can be used as proposed in [19], which is useful for both segmentation and ground filtering, but takes more time comparing to spatial segmentation. A standard (right circular) cylinder has two parallel circular ends of equal radius. In 3D, a circle is typically represented by the following parameters: the 3D location of the centre of the circle, a normal of the circle's plane, and the radius of the circle (cf., [41]). To fix a cylinder, in addition only its length has to be specified. That is, four parameters: radius (R), centre location (C), length (L) and orientation (O) completely define a standard cylinder. Fig. 2 illustrates how cylinder's centre and radius are related to circle's centre and radius considering the cylinder's ends. We first robustly fit a circle and use the estimated circle parameters: centre $c(a,b)$ and radius r for consecutive cylinder fitting. To get the parameters' set $\{R, C, L, O\}$ of a cylinder, the two new algorithms that are being proposed in this section perform the following sequential tasks (see Fig. 3).

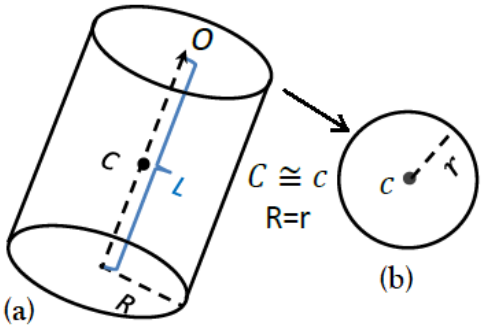

Fig. 2. Relation between cylinder and circle parameters; (a) cylinder parameters: centre C, radius R, length L and orientation O, and (b) circle parameters: centre c and radius r.

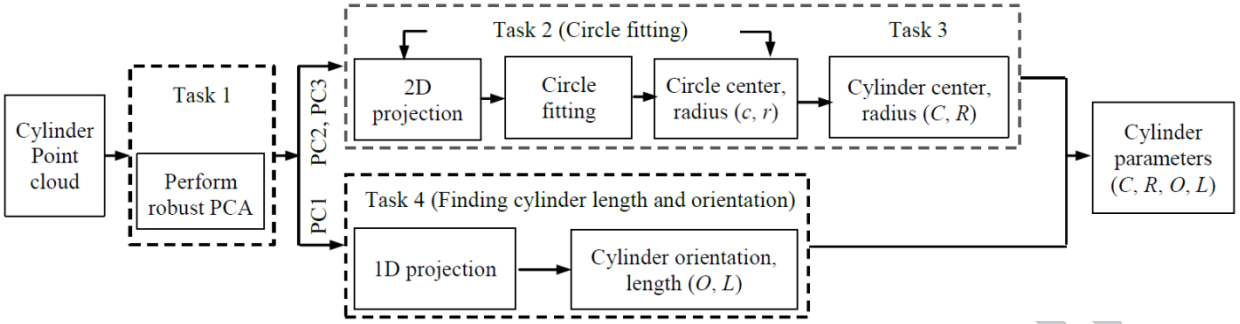

Fig. 3. Work-flow for robust cylinder fitting.

Task 1: *Performing RPCA for estimating cylinder orientation*: Recent algorithms for cylinder fitting typically employ PCA to get faster results (cf., [1, 15]). Herein, we assume that the length, L, of the cylinder is sufficiently larger than its radius, R, see Fig. 2. The PCs of the centralized variance-covariance matrix of a cylindrical point cloud indicate the data variability along and across the cylinder's axis. The PC1 shows the variability along the cylinder's axis, and the PC2 and PC3 indicate the variability perpendicular to PC1, corresponding to the width of the cylinder (see Fig. 4). Since PCA is sensitive to outliers, we employ RPCA [48] to estimate robust PCs. In this paper, the Deterministic MCD (DetMCD; [63]) algorithm is used instead of FMCD [49] to estimate robust versions of the mean and the covariance matrix in Eq. (2) to perform RPCA [48]. This is because the authors in [63] claimed that DetMCD is similarly robust as FMCD, but much faster. Hereafter, DetRPCA refers to RPCA using DetMCD based robust mean and covariance matrix. The reader is referred to Hubert et al. [63] for details on DetMCD.

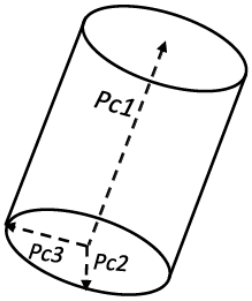

Fig. 4. Illustration of principal components (PC1, PC2 and PC3) to define the directions of a cylinder; PC1 points in general in the direction of the cylinder's axis, while PC2 and PC3 encode the data variability in the perpendicular directions corresponding to the width of the cylinder.

Task 2: *Robust circle fitting*: The radius R and centre C of the cylinder is estimated by robustly fitting a circle to projected cylinder points in this task. In Nurunnabi et al. [31], FMCD based RPCA has been used for robust circle fitting. Here for getting robust PCs as well as for increasing the overall performance of the proposed algorithms, we use DetRPCA as in Task 1. The full procedure is explained in the following three steps.

Step 1: Project 3D point cloud to get 2D points: 3D cylinder points $[p_i = (x_i, y_i, z_i)^T;\ i = 1,2,\ldots,k]$ are projected onto the plane spanned by the PC2 and PC3 that form an orthogonal basis of that plane. For this step, the PCs obtained in Task 1 are used. The principle is to project the 3D cylinder points onto the 2D robust plane. The 2D points $q_i = (q_{ix}, q_{iy})^T$ on the plane are obtained using Eq. (8)

$$q_i = (q_{ix}, q_{iy})^T = (v_o\ v_1)^T (p_i - \overline{p}). \quad (8)$$

Step 2: Estimation of robust circle centre and radius: In Step 1, the projected 2D points ideally outline a circle on the plane spanned by PC2 (v_1) and PC3 (v_o). Although, DetRPCA robustly estimates the directions of data variability, resulting 2D points are not free from outlier effects because the unordered data may contain multiple structures and we did not remove outliers. We use the Hyper [32] circle fitting method to fit a circle to the points outlying a 2D circular arc resulting from Step 1. The authors [32] showed several advantages of the Hyper algorithm, e.g., it is efficient for data sampling only part of a circle, and for circles with a smaller radius, it outperforms many existing efficient algebraic fits, it is significantly faster than the geometric fits, it has zero bias, and produces results that are invariant under rotation and translation. Fig. 5 shows however, that the Hyper algorithm fits circles imperfectly in the presence of outliers. Fig. 5 (a) shows that in case of the presence of clustered outliers, the resultant fit increases the size of the real circle (black) and consequently changes the position of the centre (blue), while Fig. 5 (b) shows that the presence of outliers in combination with an incompletely sampled circular arc biases the circle fit towards the outliers, and drastically changes its real (black) position.

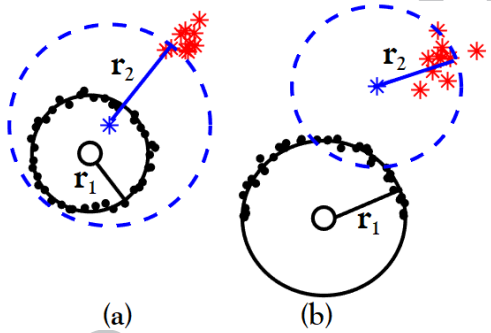

Fig. 5. Circle fitting in the presence of outliers for (a) full circle data (i.e., circle generated by the 2D projection of a full cylinder), and (b) half-circle data (i.e., circle arc generated by the 2D projection of a half cylinder). Real points and circles are in black, outliers indicated by red stars, while and circles, including circle centres, fitted using the Hyper method are in blue.

To alleviate these problems of faulty fitting because of outliers' presence, we enhance the Hyper algorithm by using the robust LTS regression principles (cf., [27]). According to LTS regression, a majority but not all observations are used for circle fitting to reduce the influence of outliers. That is, only a h-subset ($h > k/2$) of points having a minimum sum of sorted squared residuals ($min \sum_h e_i^2$) is considered for estimation. Choosing the size of h is a trade-off: the smaller h is, the more robust the results, but a larger h is expected to give more accurate estimates. To balance this trade-off between robustness and accuracy, and by incorporating the experience that usually not more than 50% outliers exist in PCD, we set $h = \lceil 0.5k \rceil$. Users can fix h based on the knowledge about the presence of outliers in their data. The algorithm starts by fitting with a randomly selected initial h_0-subset of a minimal number of points, $h_0 = 3$. This approach saves significant time for iteration because h_0 is much smaller than h. If the rank of the h_0-subset matrix is less than the dimension ($m = 3$) of the data then more points are randomly added to the h_0-subset until the rank is equal to m. A well-known Monte Carlo type probabilistic approach used also in RANSAC (cf., [20, 27]) to calculate the number of iterations I_n that is required to get at least one outlier-free minimal subset with probability p_r. Here, I_n is defined as

$$I_n = \frac{\log(1 - p_r)}{\log(1 - (1 - \epsilon)^{h_0})}, \quad (9)$$

where p_r is the probability of getting at least one outlier free subset of h_0 points and ϵ is the percentage of outliers, respectively. Based on experience, these are set at $\epsilon = 0.50$ (assuming the presence of outliers is not more than 50%), and, for getting more accuracy, $p_r = 0.999$.

Step 3: Iteration: An initial circle is fitted to the minimal h_0-subset, and the circle-parameters: centre $c_0(a_0,b_0)$, radius (r_0), and residuals e_i $(i = 1,2,...,k)$ are calculated for all points

$$e_i = \sqrt{(q_{ix} - a_0)^2 - (q_{iy} - b_0)^2} - r_0 \ . \quad (10)$$

The points are sorted in increasing order according to their squared residual, and the first *h* points having the lowest squared residuals are selected. The circle is now refitted based on these *h* points, resulting in updated centre $c_h(a_h,b_h)$ and radius (r_h) estimation.

The whole process is repeated I_n times, and the *h*-subset for which the sum of squared residuals is minimal is reported as result. The final circle parameters: centre (*a, b*) and radius *r* are the centre and radius of the circle estimated by using this final *h*-subset. Mathematically the final estimates are defined as

$$(c, r) = a\text{rgmin}\sum_{i=1}^{h} e_i^2. \quad (11)$$

This approach is named Repeated Least Trimmed Square (RLTS).

In many cases the cylinders' surface may not be smooth, e.g., when considering point clouds sampling tree trunks; many have cylindrical trunks with non-smooth bark (surface) that creates Inlier Variation (IV) in the circle arc, makes data noisy and may have loose bark that produce outliers. Fig. 6 shows a terrestrial laser scanned tree for which the surface of the trunk [Fig. 6(b)] is non-smooth resulting in inlier variations in the projected 2D points of the circular arc [see Fig. 6 (c)]. In addition, the divergence of the laser beam causes data artefacts such as angular displacement (cf., [30, 64]) which can make data noisy and non-smooth. All the above cases can create IV.

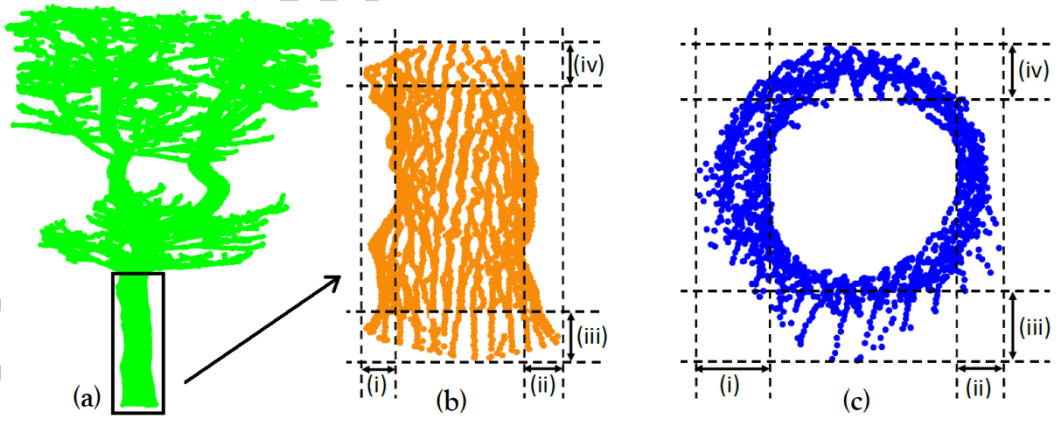

Fig. 6. Illustration of non-smooth surface resulting in inlier variation; (a) tree data, (b) tree-trunk having non-smooth surface, and (c) 2D projection of the tree-trunk onto the two PCs based plane. (i), (ii), (iii) and (iv) indicate how the projected points from Fig. 6(b) create IV on Fig. 6(c).

To cope with the influence of IV, a Weighted Least Squares (WLS) approach is developed by incorporating Tukey's well-known robust 'bi-square' weight function (cf., [65]). That is, the squared residuals of each data point are weighted according to Eq. (12)

$$w(e_i^*) = \begin{cases} \left(1 - e_i^{*2}\right)^2 for\ |e_i^*| < 1, \\ 0 \qquad for\ |e_i^*| \geq 1, \end{cases} \tag{12}$$

where $e_i^* = e_i/6\text{MAD}$ and MAD is the median of $|e_i|$. This bi-square weighting function is applied to assign less weight to points at larger distances from the considered geometric model. Cleveland [65] discussed the benefits of bi-square weighting. This weighting approach is named as Weighted RLTS (WRLTS), and the circle parameters are estimated by a small adaptation of Eq. (12) as follows

$(c, r) = a\text{rgmin}\sum_{i=1}^{k} w(e_i^*)e_i^2.$

(13)

Task 3: Cylinder's radius and centre estimation: The same radius obtained from the circle fitted in Task 2 is used as the estimated radius for the cylinder. The centre of the circle obtained in Task 2 is used to estimate a point on the cylinder axis by the following formula (cf., [15]),

$$C = \overline{p} + av_0 + bv_1, \tag{14}$$

where $\overline{p}$ is the DetMCD based robust mean of the cylinder points, and (a,b) is the centre of the circle. Note that a circle has a unique centre while the axis of a cylinder is a continuous interval of points.

This paper studies the two following special cases, which, as far we know, have not yet been considered in existing literature.

Case 1. Cylinder with different radii at its ends: such cylinders occur frequently in urban road environments, many long poles, like street light poles, have two different radii at their ends (i.e., truncated cone). This can be considered a cylinder with constantly increasing radius from top to bottom; see Fig. 7(a). In such case we estimate two different radii for their top and bottom ends. Usually, the top end of the cylindrical part has smaller radius than the bottom end. Fig. 7(b) shows that the projection of the points from the top end onto the two PCs based plane outline a smaller (inner) circle while points from the bottom end outline the larger (outer) circle. The required radii for the inner and outer circles are the distances (norm values) from the centre of the robustly fitted circle (for the complete dataset) to the point on the inner circle arc and outer circle arc, radius r_2 and radius r_1, respectively [see Fig. 7(b)]. But in reality many long poles may not be properly scanned, especially by mobile laser scanning systems because of the moving scanner. The scanned pole in Fig. 7(d) appears curved-linear rather than straight. Some of its projected 2D points (from the top end) in Fig. 7(c) create many pseudo outliers in the purple ellipse and in the yellow box because of point density variations. In the regular top view in Fig 7(e) two circles are relevant; the inner (orange) and outer (blue) circles are for the top and bottom ends of the cylindrical pole, respectively. Hence as a remedy, in such cases, we fit the inner circle by selecting points based on their z values from the top end (orange box) and fit the outer circle by taking points from the bottom end (blue box) on the pole.

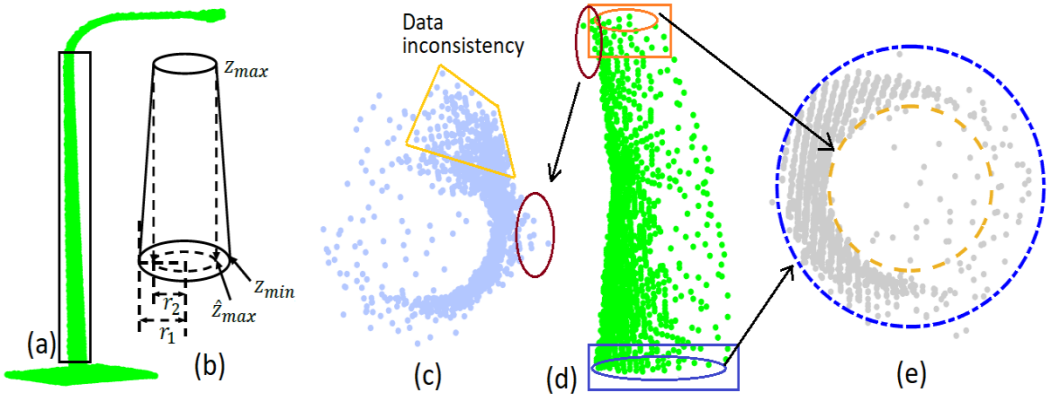

Fig. 7. (a) Street light pole; points in the black rectangle form a cylindrical part, and (b) a cylinder with different radii at the top and bottom ends; Zmin and Zmax are the minimum and maximum values of z, $\hat{Z}$max is the projected value of Zmax, r_2 and r_1 are the radii for the top and bottom ends, respectively, (c) projection of the cylinder points in Fig. 7(d), (d) cylindrical part of the pole, and (e) a regular top view of a cylindrical pole that narrows to its end.

Case 2. Cylinder with variations in data density: Such variations are typically caused by the measurement geometry, as the point density drops both with increasing range and incidence angle [66]; see Fig. 8(b). In addition, part of a cylindrical surface may not be scanned if it is at the back of a pole w.r.t. the location of the scanner, or if it is occluded by a tree branch or something else. Eq. (14) produces incorrect centre estimation if the data density varies, especially along the z-axis (cylinder's length direction). In case of point density variation in z direction, Eq. (14) estimates centre coordinates biased to the dense area. Fig. 8 shows cylindrical data with an inconsistent distribution of z values [see Fig. 8(b)]. The histogram of z-values shows such variations near the top of the pole, indicated in the red ellipse in Fig. 8(c). This inconsistency is one of the causes of an improper 2D projection resulting in a less circular shape [see in Fig. 8(d)]. To estimate correct centre for cylinders with inconsistent z-distributions, we consider the highest and the lowest values of z among the inlier values, and use their mid value rather than the average value of z in Eq. (14). The modified centre for the cylinder is defined as

$$C = \bar{\bar{p}} + av_0 + bv_1 \,, \tag{15}$$

where $\bar{\bar{p}} = (\bar{x}, \bar{y}, \frac{(z_{\max} + z_{\min})}{2})$, and $\bar{x}$ and $\bar{y}$ are the components of the DetMCD based robust mean.

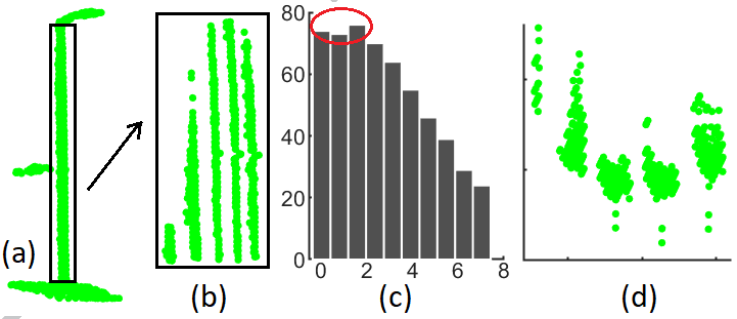

Fig. 8. (a) A street light pole, (b) zoom in of the cylindrical part, (c) histogram of z values for the cylindrical part, and (d) 2D projection of the cylinder points onto the PC2 and PC3 based plane.

Task 4: *Estimation of cylinder's orientation and length*: We use the direction of the PC1 (v_2) that is obtained by DetRPCA as orientation (O) of the cylinder. That means,

$$O = v_2. \tag{16}$$

The projection of the cylinder data onto the line spanned by the PC1 presents the cylinder axis in 1D. The i-th point (o_i) of the cylinder axis is defined as

$$o_i = v_2^T(p_i - \bar{p}). \tag{17}$$

The length of the cylinder (i.e., the length of the cylinder axis) is the distance between the maximum and minimum value of the inlier points after projection onto the cylinder axis. Therefore, the length of the cylinder L is defined as

$$L = max_i(p_i^T v_2) - min_i(p_i^T v_2), \tag{18}$$

where $p_i$belongs to the set of inlier points.

The above cylinder fitting methods based on circle-fit using Eq. (11) and Eq. (13) are named RLTS and WRLTS respectively. The proposed robust cylinder fitting process is summarized in Algorithm 1.

Algorithm 1: Robust cylinder fitting

Input:

P : Point cloud

Output:

R : radius of the cylinder
C : centre of the cylinder
L : length of the cylinder
O : orientation of the cylinder

1.	Perform DetRPCA on P and find PCs (eigenvectors v_2, v_1 and v_0), and eigenvalues λ_2, λ_1 and λ_0, where ($\lambda_2 >= \lambda_1 >= \lambda_0$).
2.	Projects all points onto v_1 and v_0 based plane so that projected points q_i are on a circle arc. Hence, $q_i = (v_0, v_1)^T(p_i - \overline{P})$,
3.	Fit robust circle on q_i, i.e., [(a, b) , r] = circle fit by RLTS or WRLTS ($q_i, h_0,\ h,\ I_n$);
4.	$C = \overline{\overline{P}} + (av_0 + bv_1)$, where $\overline{\overline{P}} = (\overline{x}, \overline{y}, \frac{z_{max} - z_{min}}{2})$
5.	$R = r$
6	$L = max_i(p_i^T v_2) - min_i(p_i^T v_2)$
7.	$O = v_2$
8.	**Output**: R, C, L, O.

3.2. Measures of evaluation

To evaluate the performance of the proposed algorithms, we estimate the cylinder parameters (C, R, L, and O), and calculate four measures: Average Distance between the real centres and the fitted centres, $\mathrm{AD}(\hat{C})$, Average estimated radius, $\mathrm{A}(\hat{R})$, Average estimated length, $\mathrm{A}(\hat{L})$, and the Average bias orientation, $\mathrm{A}(\theta)$. Bias orientation θ is defined as the angle (deviation) between the real axis and the estimated axis of the fitted cylinders. These measures are defined in Eq. (19)

$$\left.\begin{aligned} \mathrm{AD}(\hat{C}) &= \tfrac{1}{s}\textstyle\sum_{j=1}^{s}\left|C_j - \hat{C}_j\right|, \\ \mathrm{A}(\hat{R}) &= \tfrac{1}{s}\textstyle\sum_{j=1}^{s}(\hat{R}_j), \\ \mathrm{A}(\hat{L}) &= \tfrac{1}{s}\textstyle\sum_{j=1}^{s}(\hat{L}_j) \text{ and} \\ \mathrm{A}(\theta) &= \tfrac{1}{s}\textstyle\sum_{j=1}^{s}(\theta_j) \end{aligned}\right\}, \tag{19}$$

where $\hat{C}_j$, $\hat{R}_j$, $\hat{L}_j$ and θ_j are the estimates of centre, radius, length and bias orientation θ, respectively, for the j-th sample, and s is the number of samples (datasets). Bias orientation [see Fig. 9(b)], is defined as per Eq. (20)

$$\theta = \arccos\left|v_2^T \hat{v}_2\right|, \tag{20}$$

where v_2 and $\hat{v}_2$ are the real and the estimated axes (i.e. the two PC1s of the real and fitted cylinder) of the cylinder, respectively. We give priority to assess the accuracy of the positioning of the fitted cylinder above the size of the cylinder, i.e., priority of $\mathrm{AD}(\hat{C})$ and A

(θ) above $\mathrm{A}(\hat{R})$ and $\mathrm{A}(\hat{L})$. It may be the case that the estimated radius and length are equal to the real radius and length, but the cylinder is wrongly located, because the presence of similar multiple shapes is common in point cloud data. To assess the quality of the estimates, we analyse the errors of bias orientation, and calculate the Mean Squared Errors (MSE) of θ, which is defined as

$$\mathrm{MSE}(\theta) = \frac{1}{S}\sum_{j=1}^{S}[\theta_j - A(\theta)]^2, \quad (21)$$

where $A(\theta)$ is the mean (average) of θ.

4. Experiments and assessment

This section demonstrates and evaluates the proposed algorithms on both simulated and real laser scanning PCD. The algorithms are evaluated in terms of data completeness, robustness, size, consistency and inlier variation, while errors in the estimations are compared to six popular, robust and recently proposed competitive methods. Moreover, the well-known specialized software for point cloud processing, Cyclone [56], is included in the comparison considering real data experiments.

4.1. Artificial data

This section evaluates the proposed algorithms using simulated data. The advantage is that the original parameters are known. Datasets of 1000 points are generated on a straight right circular cylinder of radius R =1 meter (m) and length 10m with an axis given by the line segment of two end points (x_i, y_i, z_i); i =1 and 2. The end points are (1, 1, 1) and (1, 1, 11). The points on the cylinder surface of length (height) L and radius R are parameterized as

$$\left.\begin{array}{l} x = R\cos\varphi, \\ y = R\sin\varphi \\ z = z \end{array}\right\} \text{ and }, \quad (22)$$

where $z \in [1, 11]$ and $\varphi \in [0, 2\pi]$. To create variations among the points, random noise following a normal distribution with a Standard Deviation (StD) of 0.2m is added in all directions, i.e., to the x, y and z coordinates. We know that clustered outliers are more critical than scattered outliers to analyse. For example, the presence of clustered outliers usually causes masking and swamping errors (cf.,[27, 67]). So we include different percentages of clustered outliers in the simulations. Clustered outliers are generated with a mean location of (-2, 2, 10) and a StD of (0.3m, 0.3m, 1.5m). Fig. 9 shows an example of a simulated dataset that generates a quarter cylinder with 10% clustered outliers. Simulations are done 1000 times in each of the following experiments to get statistically significant results. In all experiments, existing algorithms are also performed with necessary parameter values as suggested in the respective papers. For the proposed RLTS and WRLTS, all parameters are set as defined in Section 3.1.

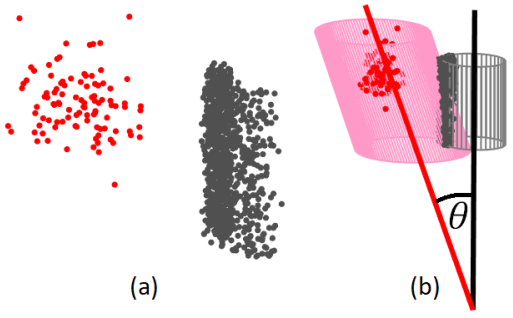

Fig. 9. (a) A simulated quarter cylinder dataset with 10% clustered outliers (red), (b) illustration of bias orientation θ, i.e., angle between the axes of the fitted cylinder (grey) without outliers and fitted cylinder (pink) with outliers. Here both the fitted cylinders are by using LS method. The black dots are cylinder points and the red points are outliers.

4.1.1. Cylinder fitting for complete and incomplete data

To explore the efficiency of the new algorithms for full and partial cylinders in the presence of outliers, we simulate datasets for full, half and one-third cylinders. For example, by a one-third cylinder we mean that points are simulated for one-third in the circular (horizontal) direction. Typically, in urban road environment, MLS captures more than one-third of the objects and it scans in horizontal direction. We generate 1000 datasets of 1000 points including 10% clustered outliers for each dataset simulating complete and incomplete sampled cylinders. We estimate cylinder parameters using the existing LS, LVH, RANSAC, MSAC, KKD, and TCL methods, and proposed RLTS and WRLTS algorithms. The evaluation measures: $\mathrm{AD}(\hat{C})$, $\mathrm{A}(\hat{R})$, $\mathrm{A}(\hat{L})$, $\mathrm{A}(\theta)$ and $\mathrm{MSE}(\theta)$ are calculated based on the estimated cylinders' parameters. The results are given in Table 1. LS and LVH perform almost similarly for full data. For incomplete data LVH fits cylinders with lower $\mathrm{AD}(\hat{C})$ and $(\hat{R})$ than LS, while reversely for bias orientation LS performs better with significantly less $\mathrm{MSE}(\theta)$. Both LS and LVH do not give acceptable results for incomplete data, even for full cylinders they estimate lengths of 13.25m and 13.18m, and produce biases in orientation, $\mathrm{A}(\theta)$, of 10.14° and 8.10°, respectively. For the same full datasets, RLTS and WRLTS produce cylinders with a length of 10.11m and 10.08m respectively, and with a small bias orientation $\mathrm{A}(\theta)$ of 0.59°. Interesting is that LS fits cylinders with less $\mathrm{A}(\theta)$ and more $\mathrm{AD}(\hat{C})$ for incomplete data compared to complete data. This is illustrated in Fig. 10(b); for a half-cylinder, LS fits a cylinder with so much bias caused by the outliers that the new fit is almost parallel but far from the real position of the simulated cylinder. On the other hand, in case of full cylinders; Fig. 10(a) shows reverse results because the fitted cylinder is biased to the outliers in its orientation but is positioned more or less correctly. That means, in the presence of outliers, the change in position is more severe than the changes in size (respect to length and radius). Existing robust methods (RANSAC and MSAC) perform significantly better than LS and LVH but produce lengths at least 2.74m larger than the real length of 10m. TCL performs slightly better than KKD, but both do not significantly improve compared to the existing robust methods RANSAC and MSAC, and are not competitive to the new algorithms. For all cases, the proposed methods using the LTS principle in an iterative fashion produce cylinders with almost the same radii and lengths of 1m and 10m respectively, with insignificant $\mathrm{AD}(\hat{C})$, while the error values of $\mathrm{A}(\theta)$ and $\mathrm{MSE}(\theta)$ are decreasing for less completely sampled cylinders.

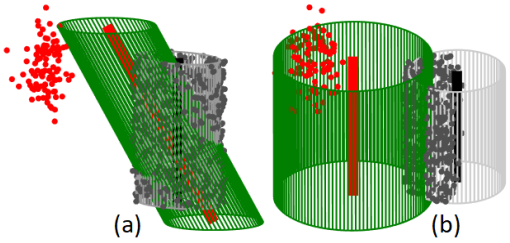

Fig. 10. Fitted cylinders by LS in the presence of 10% clustered outliers; (a) full cylinder, and (b) half-cylinder.

Table 1

Comparison of the estimated parameters ($\hat{C}$, $\hat{R}$, and $\hat{L}$ are in meter, m , and θ is in degree) for fully, half and one-third sampled cylinders.

	Full cylinder					Half cylinder					One-third cylinder				
Methods	$\mathrm{AD}(\hat{C})$	$\mathrm{A}(\hat{R})$	$\mathrm{A}(\hat{L})$	$\mathrm{A}(\theta)$	$\mathrm{MSE}(\theta)$	$\mathrm{AD}(\hat{C})$	$\mathrm{A}(\hat{R})$	$\mathrm{A}(\hat{L})$	$\mathrm{A}(\theta)$	$\mathrm{MSE}(\theta)$	$\mathrm{AD}(\hat{C})$	$\mathrm{A}(\hat{R})$	$\mathrm{A}(\hat{L})$	$\mathrm{A}(\theta)$	$\mathrm{MSE}(\theta)$
LS	0.57	1.20	13.25	10.14	103.06	2.15	1.61	12.85	1.80	4.18	2.12	1.35	12.73	0.95	1.20
LVH	0.80	1.36	13.18	8.10	65.88	1.63	1.26	13.10	6.19	38.45	1.87	1.17	12.92	5.58	31.22

RANSAC	0.80	1.00	12.78	0.81	0.95	0.87	1.01	12.88	1.10	4.56	0.92	1.03	12.74	1.17	3.58
MSAC	0.69	1.00	12.78	0.81	0.89	0.82	1.00	12.88	0.85	1.17	0.85	1.02	12.74	1.05	1.58
KKD	0.40	1.02	12.76	1.51	0.95	0.46	1.04	12.86	1.12	4.56	0.52	1.18	12.75	2.32	3.58
TCL	0.26	1.12	12.78	0.65	0.89	0.34	1.09	12.88	0.64	1.17	0.30	1.05	12.75	0.51	1.58
RLTS	0.04	1.01	10.11	0.59	0.43	0.02	1.00	10.09	0.51	0.34	0.02	1.00	10.08	0.34	0.17
WRLTS	0.02	1.00	10.08	0.59	0.43	0.02	1.00	10.09	0.51	0.34	0.05	1.00	10.10	0.34	0.17

4.1.2. Robustness and influence of outliers

To evaluate the influence of outliers on cylinder fitting, we create 1000 quarter cylinder datasets each consisting of 1000 points for three cases: without outliers, with 20% scattered outliers and with 20% clustered outliers. We fit cylinders using all the methods, and all resulting quality measures are given in Table 2. Results for data without outliers show that the non-robust methods: LS and LVH that use ordinary least squares and PCA respectively perform likewise RLTS and WRLTS, and better than other existing methods. Moreover the LS fits cylinders with the smallest values of A(θ) and MSE(θ). If we see the results for data with outliers; RANSAC and MSAC perform significantly better than LS and LVH. Especially for data with clustered outliers the differences of the results with non-robust methods are larger than for data with scattered outliers. For example, LVH and RANSAC fit cylinders for data with 20% clustered outliers having results AD($\hat{C}$) = 3.49m and 0.83m respectively, and for data with 20% scattered outliers AD($\hat{C}$) =1.84m and 0.57m, respectively. It is evident that clustered outliers have more influence on the fits than scattered outliers, and as for scattered outliers, their influence partly levels out. Fig. 11 shows that in the presence of 20% clustered outliers, many outlying points are considered as inliers by the existing methods, and results in Table 2 (Column 14) show that all existing methods overestimate the length of the cylinders by at least 3.03m (e.g., for MSAC) because their position is affected by the outliers. These results strongly indicate that even robust methods like RANSAC and MSAC are influenced by the masking effect (i.e., outliers work as inliers). Although KKD utilises RANSAC coupled with LMS it does not perform significantly better than RANSAC or MSAC. In the presence of outliers both the new methods perform almost the same and significantly better than any other existing method.

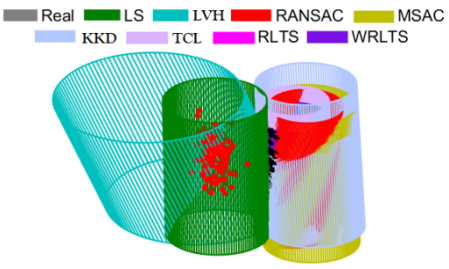

Fig. 11. Fitting results for the different considered methods in case of a quarter cylinder dataset in the presence of 20% clustered outliers. The figure shows that the newly proposed methods RLTS and WRLTS are not influenced by outliers and perform best.

Table 2

Comparison of the estimated parameters ($\hat{C}$, $\hat{R}$, and $\hat{L}$ are in meter, and θ is in degree) for data without, and with (20% scattered and 20% clustered) outliers.

	Without outliers					20% scattered outliers					20% clustered outliers				
Methods	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)
LS	0.03	1.00	10.09	0.08	0.01	1.01	2.01	10.10	2.02	5.96	2.57	2.59	13.12	1.73	4.74
LVH	0.08	1.01	10.09	0.23	0.08	1.84	2.80	10.09	0.58	0.41	3.49	2.84	13.38	9.65	93.25
RANSAC	0.82	1.18	10.07	1.41	3.06	0.57	1.17	10.09	1.24	2.14	0.83	1.01	13.13	1.17	3.49
MSAC	1.00	1.05	10.03	1.30	3.03	0.50	1.21	10.09	1.26	2.90	0.76	1.07	13.03	1.15	2.41

KKD	0.10	1.02	10.09	0.68	3.06	1.45	1.95	10.13	2.00	2.14	1.43	1.32	13.12	1.84	3.49
TCL	0.08	1.01	10.09	0.34	3.03	1.60	2.40	10.09	0.84	2.90	0.61	1.13	13.14	0.40	2.41
RLTS	0.03	1.00	10.09	0.24	0.08	0.07	1.01	10.08	0.25	0.09	0.03	1.00	10.08	0.27	0.11
WRLTS	0.03	1.00	10.09	0.24	0.08	0.06	1.00	10.09	0.25	0.09	0.03	1.00	10.08	0.27	0.11

To determine the robustness of the estimates, exploratory analysis is done by using well-known robust box-plots. Box-plots of distances between the simulated cylinders' centre and estimated cylinders' centre $D(\hat{C})$, $\hat{R}$, $\hat{L}$ and θ are drawn in Fig. 12(a), (b), (c) and (d) respectively. These plots are based on 1000 values of respective parameters from 1000 samples for different methods. In all figures, boxes for the existing methods are larger than for the newly proposed methods, that shows that the range of estimates is larger for the existing methods. Moreover, existing methods have more outlying cases (indicated by the red '+' signs above and below the boxes), which means that some of the estimates from the existing methods are extremely far from the others and not as regular as the results of the new methods. For example, in the presence of 20% clustered outliers, RANSAC in Fig. 12(a) and KKD in Fig. 12(b) produce the maximum number of outlying results for $D(\hat{C})$ and $\hat{R}$ respectively, whereas the new algorithms produce most robust results with only a few outlying cases. The lines for the median values (i.e. the red horizontal lines inside the boxes) of all the performance measures for the proposed RLTS and WRLTS are most close to the real parameter values, that is, according to their respective definitions, the median lines for $D(\hat{C})$, $\hat{R}$, $\hat{L}$ and θ should be along 0, 1, 10 and 0, respectively.

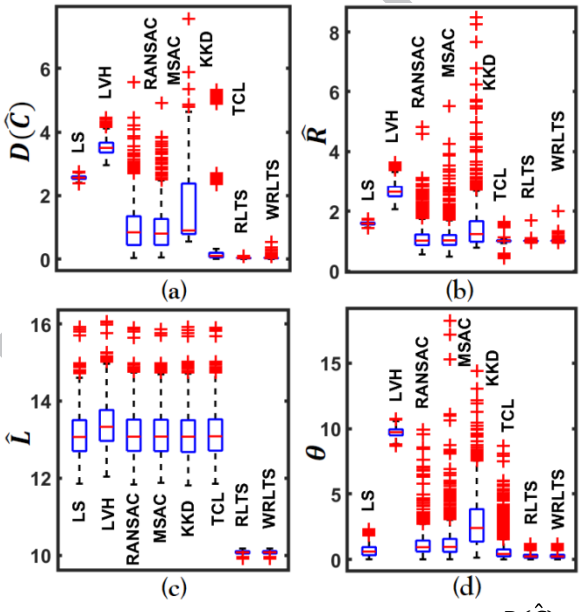

Fig. 12. Box plots of (a) distance between the centres of simulated and fitted cylinders $D(\hat{C})$, (b) estimated radii $\hat{R}$, (c) estimated lengths $\hat{L}$, and (d) bias orientations θ. Results are based on 1000 runs for quarter cylinder data with 20% clustered outliers.

4.1.3. Breakdown point evaluation

To estimate the power (level) of robustness, we calculate the robustness measure 'breakdown point' (BP; [50]). BP is estimated empirically, the parameters are calculated for the considered methods for 1000 simulated cylinders of 1000 points in the presence of 1%, 2%, …,70% scattered outliers, and 1%, 2%, …,25% clustered outliers. For each percentage of outlier contamination, $\mathrm{AD}(\hat{C})$ and $\mathrm{A}(\hat{R})$ are calculated and used for estimating BP. It is reasonable to assume that the robust methods fit cylinders at their right position and with proper size. Therefore the lines of $\mathrm{AD}(\hat{C})$ and $\mathrm{A}(\hat{R})$ should be along the virtual lines of 0 and 1, respectively, since the

robust methods are expected to produce cylinders' centres at the same centre position as used to simulate the cylinders and also reproduce the `design` radii of 1m.

Figs. 13(a) and (b) show a simulated quarter cylinder data set consisting of 1000 points including 70% scattered outliers and resulting fitted cylinders for all methods, respectively. The line diagrams, in the insets in Fig. 13(c) and (e), portray scattered outlier percentages versus $\mathrm{AD}(\hat{C})$ and $\mathrm{A}(\hat{R})$. The figures show that KKD is strongly affected by outliers. Therefore, its lines would dominate the lines from the other methods. For this reason, the KKD lines are excluded in Fig. 13(d) and (f). It is clear that even from the beginning (in the presence of only 1% outliers) existing methods fail to follow the 0 and 1 lines, while this effect increases accordingly with increasing outlier percentages. Although, in Fig. 13(d) and (f), RANSAC and MSAC perform comparatively better than LS and LVH with increasing outlier contamination, none of the results of the existing methods are satisfactory. Fig. 13(g) shows that, in the presence of clustered outliers, TCL performs best among the existing methods in case of $\mathrm{AD}(\hat{C})$. The proposed methods follow the 0 and 1 lines accordingly until they break down at 25% clustered [see Fig. 13(g) and (h)] and 70% scattered [Fig. 13(d) and (f)] outliers, respectively. So, it is justified empirically to conclude that RLTS and WRLTS have a BP of 25% and 70% for clustered and scattered outliers, respectively. On the contrary, existing methods seem always to be sensitive to outliers, even in the presence of 1% outliers. Hence it can be said that existing methods are not reasonably robust.

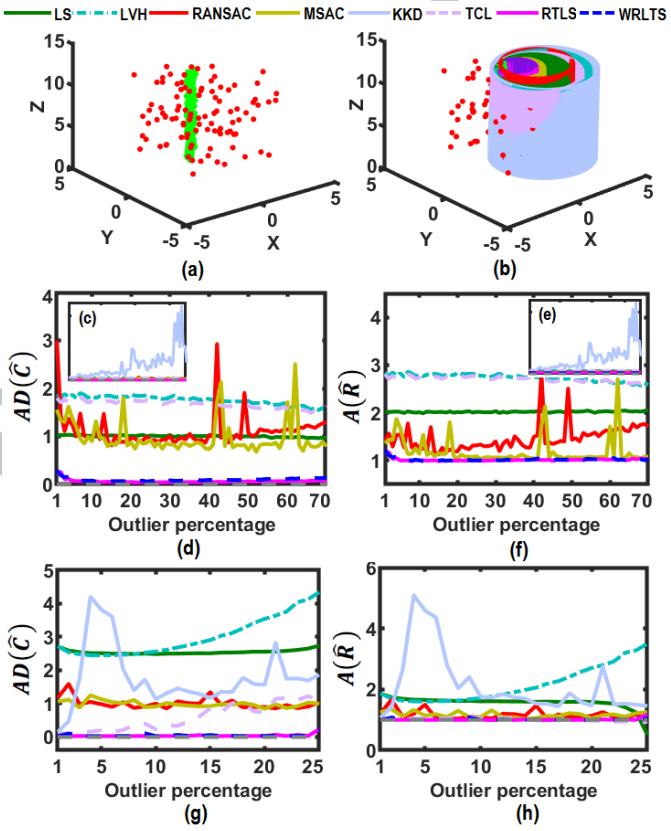

Fig. 13. (a) A simulated quarter cylinder (green) with 70% scattered outliers (red), (b) fits for the quarter cylinder with 70% scattered outliers by all concerned methods, breakdown point evaluation; line diagrams for scattered outliers' percentages 1% –70% versus (c) $\mathrm{AD}(\hat{C})$ for all methods and (d) $\mathrm{AD}(\hat{C})$ excluding KKD, (e) $\mathrm{A}(\hat{R})$ for all methods and (f) $\mathrm{A}(\hat{R})$ excluding KKD, breakdown point evaluation; line diagrams for clustered outliers percentages 1% – 25% versus (g) $\mathrm{AD}(\hat{C})$, (h) $\mathrm{A}(\hat{R})$.

4.1.4. Cylinders with different sizes

To assess the influence of the change in sizes on the fits by different methods, we generate datasets of different sizes (radius and length) of cylinder. Although we change the size, the cylinders we simulate have a fixed number of points. As the point density is the number of points per area, for a fixed number of points the variations in cylinders' sizes can also be considered as variations in point density.

To create datasets of different radii, 1000 datasets are simulated for each case with a radius of 0.50m, 1.50m and 2.00m, all of them having a fixed length of 10m. The datasets are generated as quarter cylinders consisting of 1000 points including 10% clustered outliers. The results for the different methods are given in Table 3. The table shows that LS, LVH and KKD produce significantly inferior results compared to the other techniques. The results are notably bad for the radius estimation, and large deviations for the centre locations are obtained as well as biases in the estimation of the orientation of the cylinder axis. RANSAC and MSAC perform better than LS, LVH, and KKD, and TCL produces a little better result than RANSAC and MSAC. In the case of simulated cylinders with a radius (R) of 0.50m; LS, LVH, RANSAC, MSAC, KKD, TCL, RLTS and WRLTS fit cylinders that are deviated from the real centres by 5.99m, 6.63m, 1.25m, 2.34m, 4.57m, 0.30m, 0.04m and 0.04m, respectively. TCL produces a little better result for R=0.50m and 1.50m than R=2.00m. For R=2.00m the same method fits a cylinder whose centre is 3.43m away from the real one; and produces results that are inferior in length compared to LVH, RANSAC, MSAC and KKD, and LS and LVH highly overestimate the radius at on average 40.91m and 17.52m, respectively. Most of the cases the proposed algorithms perform significantly better than existing methods, and both produce a very limited amount of errors in terms of A(θ) and MSE(θ).

Table 3

Comparison of the estimated parameters and error ($\hat{C}$, $\hat{R}$, and $\hat{L}$ are in meter, and θ is in degree) analysis for cylinders with different radii.

	R=0.50m					R=1.50m					R=2.00m				
Methods	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)
LS	5.99	5.18	12.88	12.41	163.16	3.15	1.87	12.85	2.69	7.74	43.60	40.91	7.68	16.39	665.98
LVH	6.63	6.32	13.01	6.64	44.20	2.91	1.66	12.87	4.89	24.08	16.66	17.52	8.30	2.72	11.76
RANSAC	1.25	0.87	12.75	1.81	6.36	0.90	1.50	12.71	1.22	2.48	0.72	1.51	8.22	1.12	8.45
MSAC	2.34	1.98	12.72	1.87	21.26	0.94	1.59	12.72	1.22	3.15	0.61	1.27	8.22	0.79	1.59
KKD	4.57	4.62	12.72	4.97	6.36	1.49	2.70	12.80	4.98	2.48	0.84	2.04	8.43	3.56	8.45
TCL	0.30	0.64	12.77	0.91	21.26	0.94	1.49	12.72	0.60	3.15	3.43	2.30	8.12	0.52	1.59
RLTS	0.04	0.48	10.08	0.14	0.03	0.03	1.50	10.08	0.41	0.25	0.02	2.31	9.56	0.42	0.43
WRLTS	0.04	0.48	10.08	0.14	0.03	0.08	1.51	10.14	0.41	0.25	0.05	2.31	9.61	0.42	0.43

We also generate datasets simulating cylinders of different lengths. 1000 quarter cylinders are simulated consisting of 1000 points including 10% clustered outliers for lengths of 3m, 8m and 20m, and all with 1m radius. The same methods as above are performed and used to evaluate the experiments. The results in Table 4 show that all existing methods overestimate the length by at least 2.6m for any size of length. For example, for simulated cylinders of lengths of 3m, 8m and 20m, KKD fits cylinders with lengths of 7.49m, 10.96m and 22.60m, respectively, whereas RLTS and WRLTS obtains lengths of 3.14m, 8.09m and 20.03m, respectively. Both the proposed methods have the least and very small values for MSE(θ) of 1.14, 0.16 and 0.02 for 3m, 8m and 20m long cylinders, respectively. A remarkable finding is that with a fixed radius the accuracy of the estimated length is increasing with the real length of the simulated cylinders.

These experimental results strongly indicate that the new algorithms are hardly affected by changes in the cylinders' sizes (lengths and radii) and point density.

Table 4

Comparison of the estimated parameters and error ($\hat{C}$, $\hat{R}$, and $\hat{L}$ are in meter, and θ is in degree) analysis for cylinders with different lengths.

Methods	3m					8m					20m				
	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)
LS	2.42	1.57	7.49	1.10	1.10	2.45	1.59	10.92	0.78	0.82	2.65	1.64	22.75	0.64	0.52
LVH	2.74	1.92	7.21	16.65	16.65	2.55	1.71	11.17	6.96	48.66	2.66	1.69	22.87	3.01	9.10
RANSAC	0.51	1.06	7.45	4.56	4.56	0.71	1.04	10.95	1.66	4.33	1.71	1.12	22.76	0.68	1.28
MSAC	0.43	1.11	7.41	5.97	5.97	0.82	1.12	10.92	1.83	6.31	1.68	1.11	22.76	0.70	0.96
KKD	0.23	1.01	7.49	1.57	1.57	0.61	1.25	10.96	2.70	4.33	1.64	1.83	22.60	6.99	1.28
TCL	0.41	1.10	7.49	0.74	0.74	0.38	1.08	10.94	0.45	6.31	0.56	1.13	22.77	0.45	0.96
RLTS	0.03	1.00	3.14	1.14	1.14	0.03	1.00	8.09	0.34	0.16	0.04	1.00	20.03	0.12	0.02
WRLTS	0.03	1.00	3.14	1.14	1.14	0.03	1.00	8.09	0.34	0.16	0.04	1.00	20.03	0.12	0.02

4.1.5. Consistency of the estimates

We have also assessed the effect of the number of points sampled from the simulated cylinder on the fitting, as well as the consistency among the estimated parameters. A statistically consistent estimator is one that becomes closer to the real parameters with increasing sample size [68]. To check consistency, we fit 1000 simulated quarter cylinders with a length of 10m and a radius of 1m for each of the following sample sizes, k =100, 500 and 10,000 points, including 10% clustered outliers. Different number of points for a fixed size of cylinder can also be considered as the case of varying the point density. We calculate the measures of evaluation accordingly for different methods. The results are given in Table 5.

We observe that for the existing methods, the quality of fit is not improved with an increased number of sample points, which is expected for a consistent estimator. In some cases the results are even worse when the sample size is larger, e.g., LS, LVH, RANSAC, MSAC, KKD and TCL fit cylinders with average lengths A($\hat{L}$) =11.01m, 11.20m, 11.03m, 11.00m, 10.93m and 10.92m respectively for k =100, and A($\hat{L}$) = 12.75m, 12.96m, 12.76m, 12.76m, 12.78m and 12.77m respectively for k =10,000. For the proposed algorithms, all estimates are closer to the real parameters, and so the measures AD($\hat{C}$), A($\hat{R}$) and A($\hat{L}$) indeed get better. Additionally, RLTS and WRLTS fit cylinders of A($\hat{L}$) = 9.79m and 10.08m for k=100 and 10,000 respectively, which are the closest to the real length of 10m. The error values for A(θ) = 0.43°, 0.38° and 0.28° are decreasing as their corresponding MSE(θ)= 0.82°, 0.21° and 0.11° with increasing values of k =100, 500 and 10,000, respectively.

Table 5

Influence of the number of sample points on the estimated parameters ($\hat{C}$, $\hat{R}$, and $\hat{L}$ are in meter, and θ is in degree).

	K =100					K =500					K =10,000				
Methods	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)
LS	2.51	1.62	11.01	1.82	4.75	2.50	1.62	12.37	1.05	1.66	2.50	1.62	12.75	0.89	1.13
LVH	2.56	1.70	11.20	5.54	31.39	2.57	1.72	12.57	5.64	31.99	2.56	1.71	12.96	5.73	32.95

RANSAC	0.89	1.19	11.03	1.66	5.17	0.89	1.09	12.38	1.38	3.15	0.92	1.11	12.76	1.24	3.60
MSAC	1.19	1.26	11.00	1.79	10.43	1.08	1.27	12.37	1.49	6.37	0.88	1.07	12.76	1.25	3.21
KKD	2.93	1.52	10.93	1.97	5.17	2.32	4.07	12.45	3.67	3.15	1.26	1.88	12.78	3.65	3.60
TCL	2.40	1.55	10.92	1.55	10.43	0.48	1.11	12.39	0.68	6.37	0.20	1.04	12.77	0.39	3.21
RLTS	0.12	1.03	9.79	0.43	0.82	0.04	1.01	10.03	0.38	0.21	0.03	1.00	10.08	0.28	0.11
WRLTS	0.12	1.03	9.79	0.43	0.82	0.06	1.01	10.06	0.38	0.21	0.03	1.00	10.08	0.28	0.11

4.1.6. Effect of inlier variation on the estimates

Off-surface points, noise and data artefacts (angular displacement) caused by the laser beam width are different sources for Inliers Variation (IV; cf., [30, 64]). To assess the effect of IV, variation among the inliers is created by adding random noise to the inlier points at eight different levels of StD (m): 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7 in every direction (*x, y, z*). Levels of IV are defined based on experiences with TLS and MLS PCD processing. For each level of IV, 1000 quarter cylinder datasets of 1000 points with 10% clustered outliers are simulated. We perform the same methods as above for all IVs, but for brevity of space in the table we only include outcomes for three levels of IVs (0.1m, 0.5m and 0.6m). However, all results are used as input to the line diagrams in Fig. 14(a) – (d), for IV versus the quality descriptors AD($\hat{C}$), A($\hat{R}$), A($\hat{L}$) and A(θ). From the experiments discussed earlier (the breakdown point evaluation), where a value of IV=0.2 was used; we know that existing methods do not perform well in the presence of outliers. In Table 6 and Fig. 14, the main objective is to compare results between the proposed methods (RLTS and WRLTS) in the case of increasing IV. We notice in Fig. 14(a), (b) and (d) that TCL performs significantly better than other existing methods. Results in Table 6 show that RLTS is better than WRLTS for IV=0.1m, but not for IV=0.5m and 0.6m. Fig.14 indicates that both new methods produce good results until IV =0.4, while estimates from RLTS are closer to the real parameters than the estimates from WRLTS. But for IV=0.5m, 0.6m and 0.7m, WRLTS performs significantly better than RLTS. We observe WRLTS results for IV=0.7m and beyond are not satisfactory, See Fig. 14.

Table 6

Parameters ($\hat{C}$, $\hat{R}$, and $\hat{L}$ are in meter, and θ is in degree) estimation results for different methods in case of various Inlier Variations (IV).

	IV=0.1m					IV=0.5m					IV=0.6m				
Methods	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)	AD($\hat{C}$)	A($\hat{R}$)	A($\hat{L}$)	A(θ)	MSE(θ)
LS	2.49	1.60	12.76	0.75	0.82	2.48	1.61	12.81	0.88	1.05	2.49	1.62	12.89	0.97	1.28
LVH	2.58	1.72	12.97	5.68	32.28	2.51	1.68	13.01	5.71	32.74	2.49	1.66	13.10	5.68	32.30
RANSAC	0.82	1.04	12.78	0.62	1.08	2.44	1.53	12.79	4.24	29.26	2.65	1.64	12.86	5.50	44.35
MSAC	0.84	0.99	12.78	0.58	0.68	1.84	1.53	12.78	3.31	19.06	2.73	2.14	12.82	4.23	48.73
KKD	1.57	1.99	12.74	4.54	1.08	1.27	1.26	12.83	2.13	29.26	1.44	1.32	12.91	2.23	44.35
TCL	0.36	1.07	12.78	0.25	0.68	0.43	1.05	12.82	0.96	19.06	0.63	1.06	12.92	1.48	48.73
RLTS	0.02	1.00	10.01	0.23	0.08	0.12	0.95	10.32	0.29	0.12	0.32	0.86	10.53	0.30	0.22
WRLTS	0.05	1.01	10.03	0.23	0.08	0.09	0.98	10.23	0.29	0.12	0.16	0.98	10.35	0.30	0.12

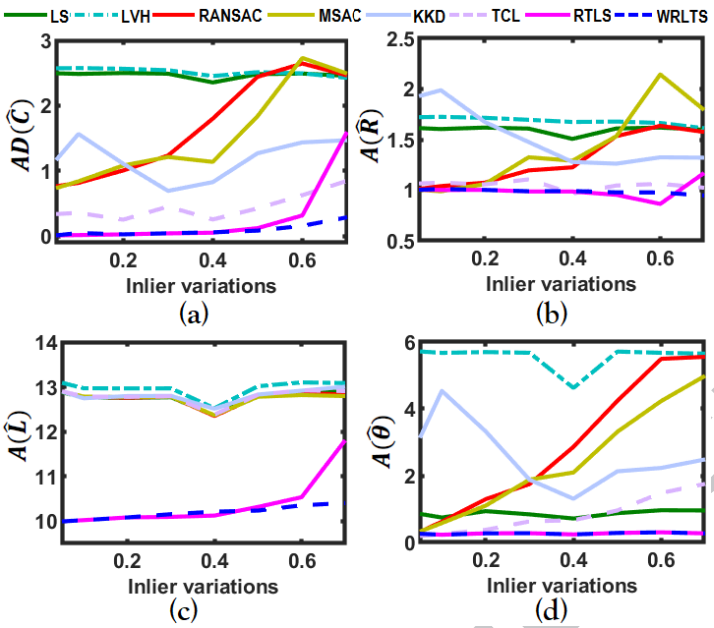

Fig. 14. Line diagrams for the relation between Inlier Variations (IV: 0.05m, 0.1m, 0.2m, 0.3m, 0.4m, 0.5m, 0.6m and 0.7m) and quality measures (a) AD$(\hat{C})$, (b) A$(\hat{R})$, (c) A$(\hat{L})$, and (d) A(θ) for all concerned methods.

4.1.7. Fitting speed

Trade-off between computation time and accuracy of the estimated parameters is a common scenario in point cloud processing. In this paper we put more emphasis on accuracy and robustness (ability to deal with the influence of outliers) of the estimates than on the time needed to acquire a cylinder fitting. In the previous sections, we have shown that the robust methods (RANSAC, MSAC, KKD, TCL, RLTS and WRLTS) produce significantly better results than the classical methods (LS and LVH) in terms of accuracy and robustness. Although the KKD was claimed to be robust, we have exposed (see Section 4.1.3) that it does not tolerate even 1% outliers, so we do not consider KKD as a robust method. In this section, we compare the computational speed only for the robust methods: RANSAC, MSAC, TCL, RLTS and WRLTS.

We evaluate the fitting speed as a function of outlier percentage, number of sample points and completeness in sampling a cylinder. To get statistically significant results, we generate 1000 cylinder (having length 10m and radius 1m) datasets of (i) one-fourth cylinders with different percentages of 5%, 15% and 25% clustered outliers with a fixed size of 1000 sample points, (ii) one-fourth cylinders of different sample sizes: 500, 1000, and 10,000 including 20% clustered outliers, and (iii) full, half and one-third cylinder data of 1000 points including 20% clustered outliers. We fit cylinders using RANSAC, MSAC, TCL, RLTS and WRLTS. Results are averaged computation times from 1000 runs for each and every sample. Results are counted in seconds (s) using the MATLAB profile function. Cylinder fitting times for different percentages of outliers, different number of sample points and completeness in sampling (full, half and one-third cylinder) are shown in Table 7 Columns 2–4, Columns 5–7, and Columns 8–10, respectively. Results in Table 7 show that in every case the proposed methods (RLTS and WRLTS) take significantly less time than the existing robust methods (RANSAC, MSAC and TCL). For example, in the presence of 25% outliers, RANSAC, MSAC and TCL fit cylinders in on average 1.637s, 1.517s, and 0.982s respectively, which are around 11, 10 and 7 times more than RLTS (0.146s). It is also observed that for the existing methods, computation time increases with increasing percentages of outliers. For example, TCL takes on average 0.982s in presence of 25% outliers, which is 2.3 times more than the time (0.426s) needed in presence of 5% outliers. That also indicates the influence of outliers' presence on the existing methods. On the other hand, the proposed algorithms,

e.g., RLTS fits cylinders with very similar times in the presence of different percentages of 5% (0.144s), 15% (0.146s) and 25% (0.146s) outliers. For a cylinder of 500 points RLTS takes 0.080s for fitting, which is around 13 and 5 times quicker than RANSAC (1.019s) and TCL (0.417s), respectively. For a one-third cylinder of 1000 points including 20% outliers, both RLTS (0.144s) and WRLTS (0.146s) are at least 5 times faster than the existing robust methods: RANSAC (0.803), MSAC (0.743s) or TCL (0.740s). It is observed that instead of using bi-square weighting as an additional task in WRLTS, the processing times for RLTS and WRLTS are not significantly different.

Although the focus of this work is on estimation accuracy we conclude that the computational performance of the introduced methodology is actually better than existing robust (RANSAC, MSAC and TCL) methodology on the same datasets, while the introduced methodology also scales better than the existing approaches w.r.t. the percentage of outliers present. An interesting upcoming application in autonomous driving is the real-time, on board matching of e.g., cylindrical objects extracted from a newly acquired point cloud to an existing 3D database of road-side objects. It is expected that the methodology as introduced in this manuscript could contribute to such application, but further testing is required to assess if the methodology can indeed be applied at real-time.

Table 7

Cylinder fitting time (in seconds).

Methods	Outlier percentage			Number of sample points			Cylinder portion		
	5	15	25	500	1000	10,000	Full	Half	One-third
RANSAC	0.167	0.568	1.637	1.019	1.143	1.207	0.560	0.652	0.803
MSAC	0.156	0.549	1.517	1.053	1.067	1.134	0.567	0.635	0.743
TCL	0.426	0.561	0.982	0.417	0.792	8.848	0.739	0.768	0.740
RLTS	0.144	0.146	0.146	0.080	0.161	0.563	0.132	0.160	0.144
WRLTS	0.145	0.146	0.149	0.081	0.164	0.572	0.133	0.164	0.146

4.2. Real datasets

This section evaluates the proposed algorithms on real laser scanning datasets. Notably results are compared to those obtained by existing methods. Additionally, in this section on real data experiments we consider the results for cylinder fitting using the Cyclone software [56].

(1) *Traffic signal pole data:* We consider a dataset sampling a roadside pole carrying a traffic signal. This dataset of 799 points is collected by a vehicle-borne mobile laser scanning (MLS) system. It has a point density of 200pt/m^2. To fit a cylinder, we select a vertical (cylindrical) portion of the pole, as indicated by the rectangle in Fig. 15(a). The length of the pole is 2.94m and its radius is 5.7cm. As usual for MLS, the pole is not fully sampled [see Fig. 15(b); side view], if we look horizontally [Fig. 15(c); top view] through the cylinder, it appears half cylindrical. Many points in the red ellipses around the top of the pole sample light bulbs and their support baskets, these are clearly pseudo outliers for the cylinder fitting.

We run all the existing methods using parameters settings as advised by the respective authors in their papers. In case of the proposed algorithms, parameters are fixed as defined in Section 3.1. To obtain estimates for the radius and centre location of the cylindrical part of the pole [Fig. 15(b)] first we fit a circle (magenta) according to the proposed RLTS algorithm for the projected 2D points [Fig. 15(d)] in the DetRPCA based plane. The figure shows that the PCA (in LVH) based circle (blue-green) is biased to the outliers and has a much bigger radius than DetRPCA based RLTS and WRLTS. Table 8 (Columns 3 and 5) gives an overview of the estimates obtained by the different methods, while Fig. 15(h) – (p) illustrate the results of the fits (grey), including Cyclone [Fig. 15(l)]. Real cylinder points are colored according to their residual distances to the model fits for the respective methods. LVH and

TCL fit cylinders similar to LS that are significantly wider than the other methods (i.e., more biased to the outliers), because both approaches use PCA in combination with algebraic fitting, which is influenced by outliers. Fig. 15(m) shows that KKD fits a cylinder with a wrong orientation: the fitted cylinder (grey) is not aligned to the real colored points. KKD fails to fit accurately since it uses LMS at its initial fitting stage with a very limited number of iterations. In Section, 2.2.2 we discussed that LMS has convergence issues. RANSAC and MSAC do not fit better than other existing methods, but MSAC (6.3cm) obtained a better radius than RANSAC (7.6cm). Cyclone estimates radius 5.7cm, but overestimates length (3.06m), which is the worst result among the other in the comparison.

To validate the estimates, we manually clean the data, the result is shown in Fig. 15(e), and the result is free from any outlier. Fig. 15(f) shows these cleaned points projected onto the PC2 and PC3 based plane. We have used LS cylinder fitting on the cleaned data and have compared the results with the results from the other methods in the presence of outliers. We consider these cleaned LS results as the benchmark because for outlier free data, LS is one of the best methods for model fitting as discussed above (see Table 2). Fig. 15(g) shows the LS fit to the clean data. It estimates the radius of the poles at 5.7cm. We also perform the other methods on the clean data; the resulting estimated radii are given in Column 2, Table 8. Comparison between Column 2 (radii for clean data) and Column 3 (radii for real data) shows the changes (influence of outliers) for the respective methods. It shows that all existing methods are influenced by outliers. RANSAC produces different radii depending on the presence of outliers, whereas MSAC produces the same radii both for real and clean data, which may be the case for using robust cost functions in its fitting. For the clean data Cyclone does better for estimating length than for the case of real data, but creates a huge gap (13cm) between the lengths for the cleaned (2.93m) and real (3.06m) data, which is worse than any other methods. Our proposed methods, implementing the LTS principle, use the majority of the real inlier surface points, which are evident in the residual plots [see Fig. 15(o) and (p)], and produce accurate values for the radius of 5.7cm and for the length 2.937m (for RLTS) of the cylinder that coincide with the values obtained by physical measurements (Ground Truth; GT: radius 5.7cm and length 2.94m) of the pole.

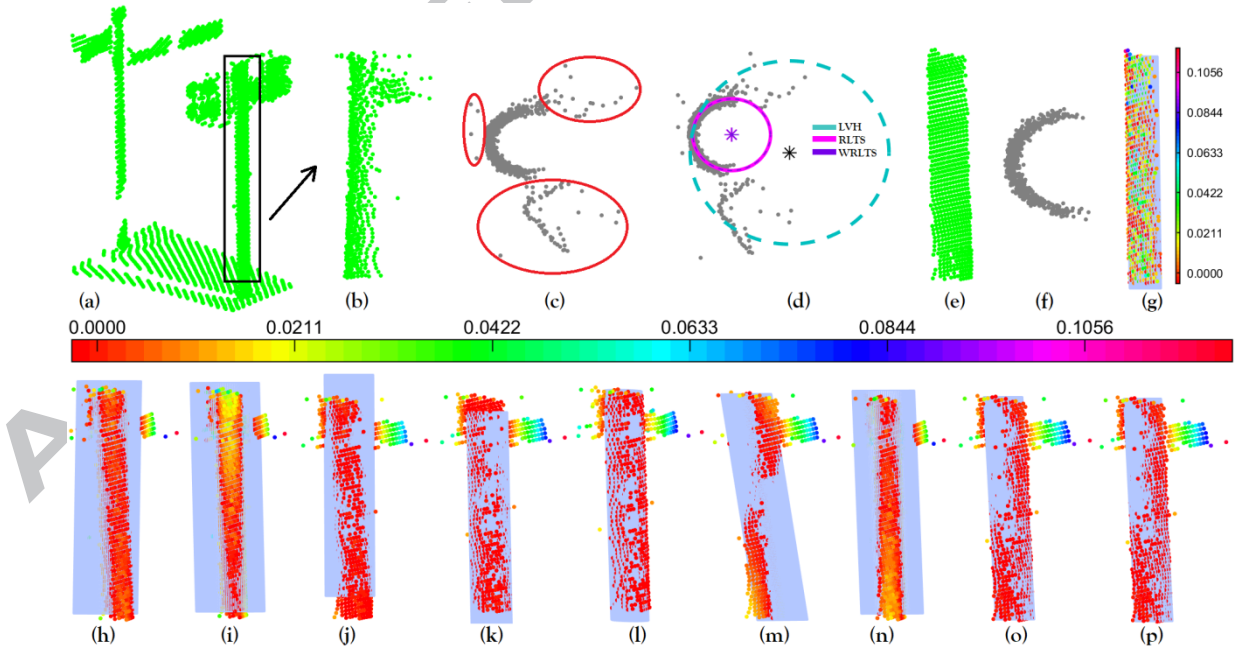

Fig. 15. Fitting a cylinder to a traffic signal pole (a) traffic pole data, (b) cylindrical data; side view, (c) cylindrical data; top view; outliers are highlighted in red ellipses, (d) projection of the sampled data onto 2D plane and fitted circles by PCA based LVH (blue), and DetRPCA based RLTS (magenta) and WRLTS, (e) clean data (f) projection of the clean data onto 2D plane, (g) LS fit and corresponding residuals of the clean data, fitted cylinders and residual distances, (h) LS, (i) LVH, (j) RANSAC, (k) MSAC, (l) Cyclone (m) KKD, (n) TCL, (o) RLTS, and (p) WRLTS.

Table 8
Estimated radius (in meter; m) and lengths (m) for the roadside traffic signal pole data.

Methods	Radius(m)		Length(m)	
	Clean data	Real data	Clean data	Real data
GT	0.057		2.940	
LS	0.057	0.112	2.937	2.952
LVH	0.057	0.147	2.937	2.954
RANSAC	0.069	0.076	2.936	2.953
MSAC	0.063	0.063	2.953	2.952
Cyclone	0.055	0.057	2.930	3.060
KKD	0.122	0.078	2.936	2.963
TCL	0.089	0.130	2.937	2.955
RLTS	0.058	0.057	2.937	2.937
WRLTS	0.057	0.057	2.935	2.933

(2) *Bollard data:* This dataset is scanned by a TLS system at the Main Roads Western Australia MLS test site. It contains 11 similar bollards (short traffic poles), has 17,427 points with a point density of 242pt/m^2, compare Fig. 16(a). We call this dataset the 'bollard' dataset.

This 'bollard' dataset needs pre-processing as it includes ground points. We go through three sequential steps: (i) ground point filtering using the LAStools software [69], (ii) vertical linear feature (pole) classification [70, 71], and (iii) spatial segmentation (region growing based on Euclidean distance with a pre-assigned threshold) to get individual pole like objects in Fig. 16(b). We filter ground points by running LAStools. Next we perform the method in [70] to calculate the Geometric Index (GI) to identify linear features, and use the following parameters: k nearest neighbourhood size, $knn_{th} = 50$; and presence of outliers' percentage, $\epsilon = 0.25$, and points are considered linear if GI > 0.6. Finally, we perform spatial segmentation using two parameters: Euclidean distance threshold, $Ed_{th} = 0.2$m and $knn_{th} = 50$ to separate the 11 bollards. The reader is referred to [70] for more details on GI, and parameters. Fig. 16(b) shows that individual bollards still contain some non-cylindrical points (outliers) from the ground and from some other parts. For example, the non-cylindrical points in the red ellipses in the top view [Fig. 16(d)] for Bollard 11 [Fig. 16(c)] appear as outliers. We perform all methods, and run Cyclone for fitting cylinders to individual bollards contaminated by outlying points. The results (mean and StD) of the 11 estimated radii and lengths for the respective techniques and software are given in Table 9. In addition, we fit cylinders using LS to the manually cleaned data, and obtained a radius of 7.3cm and a length of 55.2cm. This results for LS fit has not been included in the Table. We collected physical measurements for the respective bollards and averaged them, giving results (radius 7.3cm and length 55.0cm) similar to the values of the LS fit on the cleaned data. For the outlier contaminated raw data, all existing methods overestimate the radius, while Cyclone underestimates. For example, LS, LVH and Cyclone fit cylinders with radii of 8.4cm, 10.8cm and 7.1cm respectively, whereas the real one is 7.3cm, which is only obtained by the proposed RLTS and WRLTS methods. Most methods approximate the real lengths of the bollards in their estimates. RLTS and WRLTS estimate the lengths of 55.3cm and 54.3cm respectively, with a small variation (Std; 0.017). Fig. 16(e) shows that RLTS avoids outliers when fitting a cylindrical model to Bollard 11. On the other hand, existing methods are influenced by outliers and over fit the bollard as shown in Fig. 16(f).

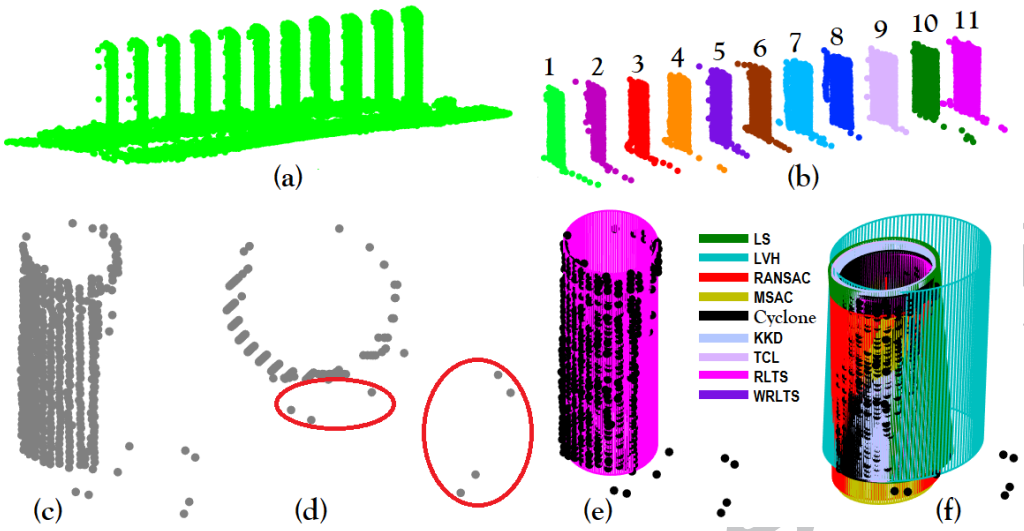

Fig. 16. Fitting cylinders to bollard dataset; (a) bollard dataset, (b) individual bollards, (c) larger view of Bollard 11, (d) top view of Bollard 11 showing outlying points in red ellipses, (e) RLTS fit to the Bollard 11, and (f) all the fits from different methods are overlaid in one figure.

Table 9

Estimated radii (m) and lengths (m) of the bollards.

Estimate	Measure	Methods									
		GT	LS	LVH	RANSAC	MSAC	Cyclone	KKD	TCL	RLTS	WRLTS
Radius (m)	Average	0.073	0.084	0.108	0.079	0.074	0.071	0.075	0.077	0.073	0.073
	StD		0.028	0.035	0.004	0.002	0.001	0.002	0.004	0.001	0.001
Length (m)	Average	0.550	0.569	0.570	0.570	0.568	0.556	0.569	0.569	0.553	0.543
	StD		0.020	0.020	0.019	0.019	0.013	0.019	0.019	0.017	0.017

(3) *Walkway data:* Fig. 17(a) presents a TLS dataset surveying a walkway towards the Architecture building at Curtin University, Australia, acquired by a Leica ScanStation C10 pulse based terrestrial scanner. It covers 108m^2 area and contains 620,273 points sampling pavement (ground), roof and 20 cylindrical beams of two types. In situ (physical) inspection shows that there are 10 (#11–#20) pillars of one group with a radius of 7cm and a length of 86cm that are positioned above another larger group of 10 (#1–#10) pillars with a radius of 20.2cm and a length of 4.10m.

As for the bollard dataset, we run LAStools for ground filtering, calculate GI to identify linear features using similar parameters and values: $knn_{th} =$ 50; and $\epsilon =$ 0.25, and points are considered linear if GI $>$ 0.9, and finally perform spatial segmentation with $Ed_{th} = 0.2$m and $knn_{th} =$ 50. Note that the used GI value is different from the earlier value, i.e., GI is data dependent. Results are for the 20 individually numbered pillars as shown in Fig. 17(b). All fitting methods and Cyclone are used to estimate their radii and lengths. Fig. 17(c; bottom) and (e) show that big pillars (#1–#10) are almost clean (without clustered outliers). Results in Table 10 show that the estimated radii of the lower parts (#1–#10) of the pillars in the walkway are reasonably good (around 20.2cm) by all methods, except for RANSAC, which overestimates a bit to 20.6cm. Fig. 17(c; top) and (d) shows that Pillar #20 contains some clustered outliers in the red ellipses. A similar situation is observed for most pillars (#11–#20) in that group. Hence, in the presence of outliers, for the smaller pillars (#11–#20) we get significantly different radii for the existing methods, e.g., LS, LVH and KKD

over fit cylinders by finding average radii of 12.5cm, 12.3cm and 12.1cm, with StD values of 0.087, 0.082 and 0.096, respectively. We manually clean Pillar #20, and obtain the following LS fitting results: a length of 86cm and a radius of 7cm which is the same as the ground truth. Cyclone remarkably overestimates the cylinder length at 96.1cm with a StD of 0.034. Both RLTS [see Fig. 17(f)] and WRLTS fit correct radii of 7cm and lengths of 86cm and 86.2cm with a minimal variation (StD) of 0.001 and 0.032 (and 0.030), respectively. The proposed methods produce significantly better results than any of the existing fitting techniques.

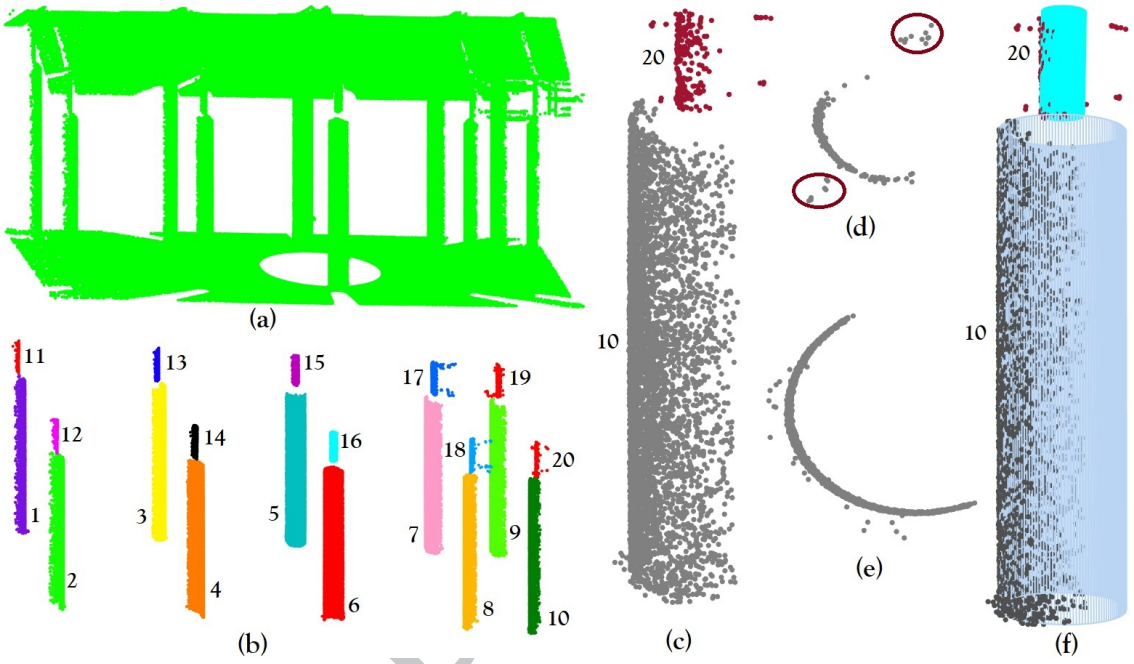

Fig. 17. Fitting cylinders to the walkway data (a) walkway dataset, (b) Twenty extracted pillars, (c) larger view of Pillars 10 and 20, (d) top view of Pillar 20; points insight the red ellipses are outliers, (e) top view of Pillar 10, and (f) RLTS fit for Pillars 10 and 20.

Table 10
Comparison for the estimated radii (m) and lengths (m) of the pillars for the Walkway data.

Estimate	Pillar No.	Measure	Methods									
			GT	LS	LVH	RANSAC	MSAC	Cyclone	KKD	TCL	RLTS	WRLTS
Radius (m)	1-10	Average	0.202	0.203	0.202	0.206	0.203	0.203	0.203	0.203	0.202	0.202
		StD		0.001	0.001	0.006	0.001	0.001	0.002	0.001	0.001	0.001
	11-20	Average	0.070	0.125	0.123	0.079	0.071	0.072	0.121	0.071	0.070	0.070
		StD		0.087	0.082	0.006	0.001	0.001	0.096	0.001	0.001	0.001
Length (m)	1-10	Average	4.100	4.127	4.128	4.128	4.127	4.130	4.128	4.128	4.100	4.108
		StD		0.024	0.024	0.024	0.024	0.043	0.024	0.024	0.009	0.017
	11-20	Average	0.860	0.878	0.879	0.879	0.879	0.961	0.806	0.878	0.860	0.862
		StD		0.033	0.033	0.033	0.033	0.034	0.193	0.033	0.032	0.030

(4) *Crossroads data:* This large MLS dataset as shown in Fig. 18(a) surveyed by an Australian local company and samples a circular crossroads in Queensland, Australia. It contains 34 different types of poles, consists of 2,436,541 points having a point density of 459pt/m^2. To justify the fitting methods for diverse sizes (lengths and radii) of cylinders, we select this highway crossroads scene, which has different sizes of cylindrical poles including 4 street lights, 19 signs, and 1 flag pole.

As for the previous real datasets, we run LAStools, calculate the GI and perform spatial segmentation. We use similar parameters and values for these tasks, the only exception being: GI $>$ 0.7. We get 34 different poles as visualized in Fig. 18(b). Poles are classified into 8 Groups (G) based on their lengths (m): G1(1–4), G2(5), G3(6–11), G4(12, 13), G5(14, 15), G6(16–21), G7(22–33) and G8(34). Finally, we perform the different fitting techniques and calculate the mean and StD of the estimated radii ($\hat{R}$) and lengths for the different methods. Since the estimations obtained by the different methods did not vary considerably; especially for the lengths, we just compare the estimated radii in Table 11. Unfortunately, ground truth values were not available for this dataset. It is however concluded from the analysis of the three previous real datasets that LS fitting on outlier free (clean) data gives results that are almost the same as ground truth (GT). We select one comparatively well scanned pole as benchmark for each of the 8 groups and clean it manually. Cylinders are fitted by LS to the resulting outlier free point cloud for the eight different groups. These clean data results for the eight groups are given in Column 2, Table 11. Pole group G7 contains 12 (Poles 22–33) poles, we manually clean Pole 27 and estimate its radius at 3.4cm. For unclean data most of the existing methods significantly overestimate this radius. Although KKD uses robust LMS, and TCL uses normal based PCA in an iterative fashion, they obtain average radii of 23.1cm and 1.05m respectively, which are highly overestimated. RANSAC, MSAC and Cyclone perform better than the other existing methods. The new algorithms estimate the radii of G7 at on average 3.2cm with very small variations (0.002, 0.003), they are the more close to the real one. This behaviour is not only observed for group G7; RLTS and WRLTS perform overall better than the existing methods for almost all groups.

An exceptional situation occurs in Group 1, consisting of Poles 1–4. Figs. 18(b) and (c) show that these long light poles have radii that increase from top to bottom, that means they have two different radii at their ends and of varying data density. The centre is estimated by Eq. (15). The proposed RLTS and WRLTS algorithms find average radii of 12.7cm and 6cm at the bottom and top ends of the poles, respectively, with a very limited StD. Whereas the existing methods, including Cyclone, find just one radius for the poles, which is obviously wrong. After manual cleaning Pole 1, LS estimates its radius at 9.7cm, this value is also closer to the value 9.4cm, which is the average of the two outer radii (12.7cm and 6cm) determined by RLTS; it indicates that the two radii at the ends of the pole are properly estimated. Similar successful evidence is found for WRLTS.

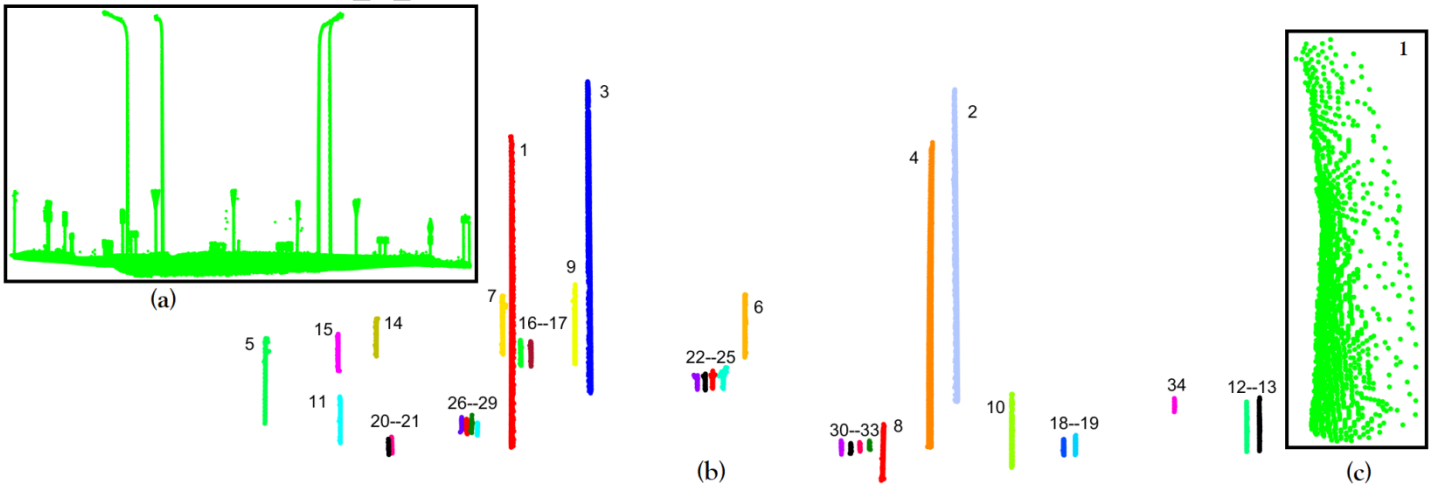

Fig. 18. Fitting cylinders to the crossroads dataset (a) crossroads data, (b) extracted cylindrical poles, and (c) a prototype of Group 1 (Poles 1-4); Pole 1.

Table 11
Radii (m) estimated by the different methods for the crossroads data.

Group (Pole)	Clean data	Real data									
		Methods									
	LS	Results	LS	LVH	RANSAC	MSAC	Cyclone	KKD	TCL	RLTS	WRLTS
G1 (1-4)	0.097	$A(\hat{R})$	0.098	0.101	0.102	0.101	0.099	0.102	0.111	0.127,0.060	0.127,0.060
		$StD(\hat{R})$	0.005	0.005	0.006	0.014	0.003	0.022	0.012	0.003,0.007	0.003,0.004
G2 (5)	0.039	$\hat{R}$	0.078	0.094	0.061	0.043	0.038	3.104	0.112	0.039	0.039
G3 (6-11)	0.039	$A(\hat{R})$	0.047	0.058	0.042	0.038	0.032	0.076	0.066	0.034	0.033
		$StD(\hat{R})$	0.012	0.020	0.003	0.004	0.007	0.080	0.051	0.002	0.002
G4 (12,13)	0.034	$A(\hat{R})$	0.034	0.036	0.033	0.033	0.030	0.050	0.078	0.036	0.035
		$StD(\hat{R})$	0.000	0.000	0.001	0.002	0.005	0.003	0.055	0.000	0.000
G5 (14,15)	0.031	$A(\hat{R})$	0.046	0.051	0.049	0.047	0.037	0.024	0.046	0.037	0.037
		$StD(\hat{R})$	0.007	0.010	0.000	0.007	0.005	0.009	0.000	0.006	0.006
G6 (16-21)	0.034	$A(\hat{R})$	0.058	0.047	0.053	0.034	0.038	1.454	3.555	0.032	0.030
		$StD(\hat{R})$	0.055	0.023	0.045	0.005	0.008	1.416	6.656	0.006	0.005
G7 (22-33)	0.034	$A(\hat{R})$	0.051	0.056	0.039	0.031	0.037	0.231	1.050	0.032	0.032
		$StD(\hat{R})$	0.031	0.038	0.007	0.004	0.006	0.331	2.109	0.002	0.003
G8(34)	0.031	$\hat{R}$	0.031	0.031	0.033	0.033	0.036	0.589	0.313	0.032	0.032

5. Discussion

This section briefly discusses how and why the proposed algorithms successfully fit cylinders in 3D PCD, and gives more details on the experiments performed. In the previous section, besides simulated data, we analysed four sets of real laser scanning point clouds. For all the datasets, the new algorithms perform two major tasks for robust cylinder fitting: (i) cylinder's radius estimation, and (ii) length and orientation estimation. First, robust circle fitting is applied to estimate a cylinder's radius. Projecting 3D PCD onto the 2D plane spanned by the DetRPCA based PC2 and PC3 that creates a 2D circular arc. For every real laser scanning cylindrical dataset (1–4) the obtained circular arc was incomplete and always included some scattered and/or clustered outliers [e.g., see Fig. 17(d)]. We fitted circles to the 2D circular arcs to estimate the cylinders' radii. Here, the Hyper circle fitting method gives more reliable results in case of fitting circles to incomplete and smaller arcs. Simulation results in Fig. 5 showed that for both full and partial cylinders, Hyper could not tolerate the presence of outliers as it biases the size and position towards the outliers. Similar results were found and shown in Figs. 10 and 11. Although the Hyper method doesn't tolerate the presence of outliers, the results [e.g., see Fig. 15(d)] showed that the employment of a robust LTS regression principle, basically using a majority portion of a selected outlier free robust sample of 2D data that had a minimal sum of sorted squared residuals, successfully avoided the effect of outliers on the resultant fits, and perfectly fitted circles to estimate the radius of a cylinder. In Section 3.1 (Task 2), we explained how using LTS regression a number of times [I_n in Eq. (9); estimated by a Monte Carlo type probabilistic approach] results in an outlier free robust sample as required for RLTS and WRLTS. Experiments on the traffic pole dataset showed that none of the existing methods is free from outlier effects as all of these methods overestimate length and radius: LVH, RANSAC and TCL fitted cylinders of radii 14.7cm, 7.6cm and 13.0cm respectively. The difference between LVH and TCL is much less as they both use PCA, MSAC fitted (6.3cm) better than other existing methods as it uses a robust M-estimator [50] for its cost function as an alternative to the LS cost function used in RANSAC, whereas RLTS and WRLTS estimated radii of 5.7cm which is similar to the ground truth. Residual plots [Fig. 15(h–p)] clearly showed that all fits by existing methods were biased to the outliers, whereas the newly developed methods were free from outlier effects, mainly, because using the LTS principle, the *k*-*h* larger residuals were truncated at each

iteration. In Section 4.1.4, results on the synthetic data make it evident that in the presence of outliers, the lengths of the fits are overestimated more strongly in cases of smaller cylinders, a finding that is confirmed in Table 10 for the real Walkway dataset. Figs. 16(f) and 17(f) visually evidenced the accuracy of cylinders' parameter (radius and length) estimation for the bollard and walkway datasets respectively.

In the case of three real point clouds experiments (2, 3 and 4) that required pre-processing to get labelled (cylindrical) data, three common pre-processing steps are performed: ground surface filtering, pole like linear feature extraction and Euclidean distance based spatial segmentation. In general, methods that need pre-processing depend heavily on the success of the pre-processing methods. For example, using a segmentation method usually results in either over or under segmentation errors. For the bollard dataset in Fig. 16 (b) some outlying points were visible near each bollard, mainly because of filtering errors and under segmentation. Results in Table 9 showed that most of existing methods including RANSAC, MSAC and Cyclone, were affected by outliers; outliers appear to them as inliers, which is the well-known masking effect, but the proposed algorithms perform efficiently in the presence of under segmented outlying points. Fig. 16(f) compared the results. Moreover, the estimated StDs of the respective fits in Tables 9, 10 and 11 gave evidence that the proposed algorithms produced minimal error almost in all explored fitting scenario's.

The proposed algorithms use the DetRPCA approach to assess the directions of the data variability along and across the axes of a cylinder, while the Hyper method is used for fitting a circle invariant from rotation and translation. As such, the performance of the proposed algorithms is independent of any particular orientation of a sampled cylinder in a scene. For the same reason the positions we defined for the simulated cylinders were not dependant on any specific situation. We explored on some real data that the proposed algorithms successfully fit non-vertical cylinders in the presence of outliers, and produce significantly better results than existing methods. In this paper we concentrate mainly on fitting cylindrical objects in the road environment, hence we considered vertical poles in our examples.

In addition to the introduction of robust cylinder fitting methodology for point cloud data contaminated by outliers sampling partial and complete cylinders, this paper developed solutions for three additional cases: (i) more inlier variation, i.e., more noisy data, (ii) cylinders with different radii at their ends, and (iii) cylinder with varying data density, as the points density in real data sets typically decreases with increasing range and incidence angle. To manage more inlier variation, squared residuals in RLTS [see Eq. (13)] were weighted according to Tukey's bi-square weight function in Eq. (12). Results in Table 6 on artificial data justified that using bi-square weights in WRLTS algorithm performs significantly better than other methods including RLTS, for data with more inlier variation. In the case of poles with varying data (point) density and different radii at their ends, we could demonstrate on, e.g., the crossroad data, that the proposed algorithms estimate different radii accurately at the ends of the poles. And in the case of long poles with equal radii at the ends, averaging two radii at the top and bottom ends successfully estimate the radius of the pole. One more remark on the new algorithms, in simulation, we get a BP of 25% and 70% for clustered and scattered (uniform) outliers respectively. The reason for less BP for the presence of clustered outliers is that clustered outliers have a stronger impact than scattered outliers as the clustered outliers work together as a group to bias an estimator towards them and it is known that they can create masking and swamping effects in their own groups as well as to the inliers. It is a challenge to achieve a higher BP, at least up to 50%, in the case of clustered outliers is present in point clouds.

6. Conclusions

In this paper, two variants of robust cylinder fitting algorithms are proposed, namely RLTS and WRLTS. The main contribution is that in the presence of a high percentage of scattered (around 70%), as well as of clustered (around 25%) outliers, the new methods are applicable to both incomplete and full cylindrical point cloud data, acquired by TLS and MLS. Experiments on simulated and real laser scanning data reveal that, classical approaches: LS and PCA perform better than other methods without outliers, but in the presence of outliers, they produce misleading non-robust estimates. For RANSAC and MSAC, which are usually acknowledged as robust methods in the presence of outliers, our results show that they are also affected by outliers; such that, they overestimate the cylinder parameters (length and radius) because of biasness to outliers that make them appear as inliers (i.e. well known masking effect). In most cases, TCL [12] performs better than other existing methods, but in the presence of clustered outliers not sufficiently reliable for incomplete data. In addition, many parameters need to be tuned for PCA based TCL [12] and RANSAC based KKD [40] in order to get better results. On the contrary, the new methods use robust statistical approaches: robust regression and DetMCD based robust PCA, as an alternative to LS and PCA respectively, together with the Hyper circle fitting algorithm that is efficient for 2D partial circular data. The combination of robust variants of PCA and robust LTS regression helps to get robust PCs, while Hyper helps for fitting circles in case of partially sampled circles. Altogether we were able to present statistically robust cylinder fitting methods that are successful in the following scenarios: (i) statistically consistent, (ii) robust with varying sizes of cylinders, (iii) in case of different radii at the cylinder ends, (iv) in case of heterogeneous point density, (v) free from the effects of over and/or under segmentation errors if segmentation is used in pre-processing to get cylindrical data, and (vi) changing inlier variations. WRLTS is more efficient than RLTS in case of more inlier variation but less accurate in case of lower inlier variation. The newly proposed methods produce significantly better results than the existing methods: LS, LVH, RANSAC, MSAC, KKD, and TCL, and Cyclone software. For example, one outcome based on 1000 simulated datasets shows that in case of 10m long one-third cylindrical data with 10% clustered outliers, RANSAC and MSAC both overestimate the real cylinder length at on average 12.74m with $\mathrm{MSE}(\theta) = 3.58^{o}$ and 1.58^{o} respectively, whereas the proposed RLTS and WRLTS fit almost correctly at 10.08m and 10.10m lengths respectively. In addition, the new methods produce a very small error of $\mathrm{MSE}\ (\theta) = 0.17^{o}$. The methods are also significantly quicker than other existing robust methods like RANSAC, MSAC and TCL. The new methods have potential in many applications including structural engineering, as-built-inspection, industrial site monitoring, and forest inventory.

For the downside, in case of raw point clouds, the proposed algorithms require some pre-processing in order to get labelled cylindrical data, and are not self-sufficient for extracting multiple cylinders in complex environments like industrial plants. Further studies will consider how these algorithms can be incorporated in a workflow to be more automated for the processing of large-scale raw data sampling complex scenes with multiple cylinders. In this paper more emphasis was put on accuracy and robustness rather than processing time, next time importance will be given to achieve accuracy as well as real time processing.

Acknowledgments

This study has been carried out as a postdoctoral research supported by the Japan Society for the Promotion of Science (JSPS), whose activities are performed under The Ministry of Education, Culture, Sports, Science and Technology, and is also supported by JSPS KAKENHI Grant number 16F16010.

References

[1] 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, 67– 81.

[2] Vosselman, G., Gorte, B. G. H., Sithole, G., Rabbani, T., 2004. Recognizing structure in laser scanner point clouds. In: Int. Arch. of the Photogramm. Remote Sens. and Spatial Inf. Sci., vol. 36, no. 8/W2, pp. 33–38.

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

[4] Rabbani, T., Dijkman, S., Heuvel, F. V. D, Vosselman, G., 2007. An Integrated Approach for modelling and Global Registration of Point Clouds, ISPRS J. of Photogramm. and Remote Sens. 61, 355–370.

[5] Kühn, O., Linß, G., Töpfer, S., Nehse, U., 2007. Robust and accurate fitting of geometrical primitives to image data of microstructures. Measurement 40 (2), 129–144.

[6] Rabbani, T., Heuvel, F. V. D., 2005. Efficient Hough transform for automatic detection of cylinders in point clouds. In: proc. ISPRS Workshop on Laser Scanning, WG III/3-4, V/3, pp. 60–65.

[7] Gosliga, R. Van, Lindenbergh, R., Pfeifer, N., 2006. Deformation analysis of a bored tunnel by means of terrestrial laser scanning. In: Proc. Int. Arch. of Photogramm. Remote Sens. Spatial Inf. Sci., vol. XXXVI/5, Dresden, 25–27 September.

[8] 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), 1242–1259.

[9] Hinzen, K., Cucci, L., Tertulliani, A., 2013. Rotation of Objects during the 2009 L'Aquila Earthquake Analyzed with 3D Laser Scans and Discrete-Element Models. Seismological Research Letters 849(5), 745-751.

[10] Winkelbach, S., Westphal, R., Goesling, T., 2003. Pose estimation of cylinder fragments for semi-automatic bone fracture reduction. In: B. Michaelis and G. Krell (Eds.), Pattern recognition (DAGM 2003), Lecture notes in computer science, vol. 2781, pp. 566–573, Springer-Verlag, Heidelberg, Berlin.

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

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

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

[14] Chan, T. O., Lichti, D. D., Belton, D., 2015. A rigorous cylinder-based self-calibration approach for terrestrial laser scanners. ISPRS J. Photogramm. and Remote Sens. 99, 84–99.

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

[16] Liang, X., Litkey, P., Hyyppa, J., Kaartinen, H., Vastaranta, M., Holopainen, M., 2012. Automatic stem mapping using single-scan terrestrial laser scanning. IEEE Transactions on Geoscience and Remote Sensing 50, 661–670.

[17] Wang, J., Lindenbergh, R., Menenti, M., 2017. SigVox- A 3D feature matching algorithm for automatic street object recognition in mobile laser scanning point clouds. ISPRS J. Photogramm. and Remote Sens. 128, 111–129.

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

[19] 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), 4790–4805.

[20] Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. of the ACM 24, 381–395.

[21] Bolles, R. C., Fischler, M. A., 1981. A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In: IJCAI, Int. Joint Conf. on Artificial Intelligence, Vancouver, Canada, pp. 637–643.

[22] Lukács, G., Martin, R., Marshall, D., 1998. Faithful least-squares fitting of spheres, cylinders, cones and tori for reliable segmentation. In: European conference on computer vision, Springer, Berlin, Heidelberg, pp. 671–686.

[23] Lee, J., Kim, C., Son, H., Kim, C., 2012. Automated pipeline extraction for modeling from laser scanned data. In: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, Vilnius Gediminas Technical University, Department of Construction Economics & Property, Vol. 29, p. 1.

[24] Faber, P., Fisher, R. B., 2001. A buyer's guide to Euclidean elliptical cylindrical and conical surface fitting. In: Proc. BMVC, Manchester, September, pp. 521–530.

[25] Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004. Assessing forest metrics with a ground-based scanning lidar. Canadian Journal of Forest Research 34(3), 573–583.

[26] Wang, H., Suter, D., 2003. Using symmetry in robust model fitting. Pattern Recognit. Lett. 24(16), 2953–2966.

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

[28] Su, M. C., Chou, C. H., 2001. A modified version of the K-means algorithm with a distance based on cluster symmetry. IEEE Trans. on Pattern Anal. and Mach. Intell. 23(6), 674–680.

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

[30] Forsman, M., Börlin, N., Olofsson, K., Reese, H., Holmgren, J., 2018. Bias of cylinder diameter estimation from ground-based laser scanners with different beam widths: A simulation study. ISPRS Journal of Photogrammetry and Remote Sensing 135, 84–92.

[31] Nurunnabi, A., Sadahiro, Y., Lindenbergh, R., 2017. Robust cylinder fitting in three-dimensional point cloud data. In: proc. The Int. Arch. of the Photogramm. Remote Sens. and Spatial Info. Sci., vol. XLII-1/W1.

[32] Al-Sharadqah, Chernov, N., 2009. Error analysis for circle fitting algorithms. Electron. J. Stat. 3, 886–911.

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

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

[35] Chaperon, T., Goulette, F., 2010. Extracting cylinders in full 3D data using a random sampling method and the Gaussian image. In Proc. of the Vision Modeling and Visualization Conference, Stuttgart, Germany, 21–23 November, pp. 35–42.

[36] Nievergelt, Y., 2013. Fitting cylinders to data. Journal of Computational and Applied Mathematics 239, 250–269.

[37] Björck, Å., 1996. Numerical methods for least squares problems, SIAM.

[38] Awange, J. L., Paláncz, B., Lewis, R., Lovas, T., Heck, B., Fukuda, Y., 2016. An algebraic solution of maximum likelihood function in case of Gaussian mixture distribution. Australian Journal of Earth Sciences 63(2), 193–203.

[39] Paláncz, B., Awange, J. L., Somogyi, A., Rehány, N., Lovas, T., Molnár, B., Fukuda, Y., 2016. A robust cylindrical fitting to point cloud data. An International Geoscience Journal of the Geological Society of Australia 63(5), 665–673.

[40] Kawashima, K., Kanai, S., Date, H., 2014. As-built modeling of piping system from terrestrial laser-scanned point clouds using normal-based region growing. J. of Computational Design and Engineering 1(1), 13–26.

[41] 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), 633–641.

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

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

[44] Al-Subaihi, A., 2017. Orthogonal least squares fitting with cylinders. International Journal of Computer Mathematics 94(4).

[45] Joliffe, I. T., 1986. Principal Component Analysis. Springer, NY, USA.

[46] Hoppe, H., De Rose, T., Duchamp, T., 1992. Surface reconstruction from unorganized points. ACM Trans. Comput. Graphics 26(2), 71–78.

[47] De la Torre, F., Black, M. J., 2001. Robust principal component analysis for computer vision. In: Proc. 8th IEEE Intl Conf. on Computer Vision, pp. 362–369.

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

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

[50] Huber, P.J., 1981. Robust Statistics. John Wiley, New York, USA.

[51] Maronna, R.A., Martin, R. D., Yohai, V. J., 2006.Robust Statistics. Wiley, New York, USA.

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

[53] Zhang, Z., 1997. Parameter estimation techniques: A tutorial with application to conic fitting. Image Vision Comput. 15, 59–76.

[54] Hubert, M., Rousseeuw, P. J., Aelst, S. V., 2008. High–breakdown robust multivariate methods. Statistical Science 23(1), 92–119.

[55] Torr P. H. S., Zisserman, A., 2000. MLESAC: a new robust estimator with application to estimating image geometry. J. Comput. Vision Image Understanding 78(1), 138–156.

[56] Cyclone, https://leica-geosystems.com/en-US/products/laser-scanners/software/leica-cyclone; Accessed 7-01-2019.

[57] Yan, W. Y., Morsy, S., Shaker, A., Tulloch, M., 2016. Automatic extraction of highway light poles and towers from mobile LiDAR data. Optics and Laser Technology 77, 162–168.

[58] Wang, D., Hollaus, M., Puttonen, E., Pfeifer, N., 2016. Fast and robust stem reconstruction in complex environments using terrestrial laser scanning data. In: Int. Arch. of the Photogramm. Remote Sens. and Spatial Inf. Sci., vol. XLI-B3, pp.411–417.

[59] Choi, S., Kim, T., Yu, W., 2009. Performance evaluation of RANSAC family. In: Proc. of British Machine Vision Conference, London, UK, 7–10 September, pp. 1–12.

[60] Vosselman, G., Klein, R., 2010. Visualization and structuring of point clouds. In: G. Vosselman and H.-G. Maas, (Eds.), Airborne and Terrestrial Laser Scanning, Whittles Publishing, Scotland, UK.

[61] Pratt, V., 1987. Direct least-squares fitting of algebraic surfaces. Computer Graphics 21(4).

[62] Lindenbergh, R. C., Sirmacek, B., Herrero-Huerta, M., Wang, J., Berthold, D., Ebersbach, D., 2015. Automated large scale parameter extraction of road-side trees sampled by a laser mobile mapping system. ISPRS Annals II (3-W5), pp. 137–144.

[63] Hubert, M., Rousseeuw, P. J., Verdonck, T., 2012. A deterministic algorithm for robust scatter and location. J. Comput. Graphical Statist. 21(3), 618–637.

[64] El-Halawany, Lichti, D. D., 2011. Detection of Road Poles from Mobile Terrestrial Laser Scanner Point Cloud. In: Proc. Int. Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), Xiamen, 10–12 January, pp. 1–6.

[65] Cleveland, W. S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Amer. Stat. Assoc. 4(368), 829–836.

[66] Soudarissanane, S., Lindenbergh, R. C., Menenti, M., Teunissen, P. J. G., 2011. Scanning geometry: Influencing factor on the quality of terrestrial laser scanning points. ISPRS J. Photogramm. and Remote Sens. 66(4), 389–399.

[67] Nurunnabi, A. A. M., Hadi, A. S., Imon, A. H. M. R., 2014. Procedures for the identification of multiple influential observations in linear regression. Journal of Applied Statistics 41(6), 1315–1331.

[68] Lehmann, L., Casella, G., 1998. Theory of point estimation. Springer, New York, USA.

[69] Isenburg, M., Liu, Y., Shewchuk, J., Snoeyink, J., Thirion, T., 2006. Generating Raster DEM from Mass Points via TIN Streaming. In: Proc. of GIScience 2006 Conference, pp. 186–198.

[70] Rodríguez-Cuenca, García-Cortés, S., Ordóñez, C., Alonso, M. C., 2015. Automatic detection and classification of pole-like objects in urban point cloud data using an anomaly detection algorithm. Remote Sensing 7 (10), 12680–12703.

[71] Demantk´e, J., Mallet, C., David, N., Vallet, B., 2011. Dimensionality based scale selection in 3d lidar point clouds. In: Proc. Int. Arch. of Photogramm. Remote Sens. and Spatial Inf. Sci. 38(5).

Robust cylinder fitting in laser scanning point cloud data

Highlights:

- Two robust cylinder fitting algorithms are devised in laser scanning point clouds.
- The new methods fit robust cylinder in the presence of high percentage of outliers.
- The methods reliably fit partially and fully scanned cylinders.
- The proposed methods are efficient for various sizes of cylinder fitting.
- The developed methods can fit cylinders with unequal radii at their ends.