Robust Locally Weighted Regression Techniques for Ground Surface Points Filtering in Mobile Laser Scanning 3D Point Cloud Data

Abdul Nurunnabi, Geoff West, and David Belton

Abstract—This paper introduces robust algorithms for extracting the ground points in laser scanning 3D point cloud data. Global polynomial functions have been used for filtering algorithms for point cloud data; however, it is not suitable as it may lead to bias for the filtering algorithms and can cause misclassification errors when many different objects are present. In this paper robust statistical approaches are coupled with locally weighted 2D regression that fits without any predefined global function for the variables of interest. Algorithms are performed iteratively on 2D profiles: x-z and y-z. The z (elevation) values are robustly down weighted based on the residuals for the fitted points. The new set of down-weighted z values along with corresponding x (or y) values are used to get a new fit for the lower surface level. The process of fitting and down-weighting continues until the difference between two consecutive fits is insignificant. The final fit is the required ground level and the ground surface points are those that fall within the ground level and the level after adding some threshold value with the ground level for z values. Experimental results are compared with recently proposed segmentation methods through simulated and real mobile laser scanning point clouds from urban areas that include many objects that appear in road scenes such as short walls, large buildings, electric poles, sign posts and cars. Results show the proposed robust method efficiently extracts ground surface points with better than 97% accuracy.

Index Terms—Classification, Digital Elevation Model (DEM), Digital Terrain Model (DTM), ground surface extraction, Mobile Laser Scanning (MLS), outlier, robust regression, segmentation.

A. Nurunnabi is with the Department of Spatial Sciences and Cooperative Research Centre for Spatial Information, Curtin University, Perth, Western Australia 6845, Australia (e-mail: abdul.nurunnabi@ postgrad.curtin.edu.au).

G. West is with the Department of Spatial Sciences and Cooperative Research Centre for Spatial Information, Curtin University, Perth, Western Australia 6845, Australia (e-mail: g.west@curtin.edu.au).

D. Belton is with the Department of Spatial Sciences, Curtin University, Perth, Western Australia 6845, Australia (e-mail: d.belton@curtin.edu.au).

I. INTRODUCTION

In many point cloud processing tasks including corridor mapping, road assets management, infrastructure planning, environmental risk management, vegetation analysis and maintenance of urban street scenes, it is useful to classify points into ground (terrain) and non-ground points (off-terrain) [1]−[4]. Removing the ground points can simplify and minimize the time and cost of further analysis such as segmentation, feature extraction, surface reconstruction and modelling of above ground features. The same argument is valid when the objective is to get information about ground surface objects such as the road, kerb, footpath and road markings. To classify point clouds into ground and non-ground many methods have been proposed in domains such as statistics, computer vision, robotics, photogrammetry and remote sensing [2]-[8]. A comparison of the many methods was directed by the International Society for Photogrammetry and Remote Sensing (ISPRS) Working Group (WGIII/3; [8]), which showed that no method is sufficiently good for every dataset, and identified a number of issues [2]. For example, the ground cannot be characterized by geometric properties only. Looking at a local neighbourhood, similar characteristics may be encountered when comparing small terrain features and small objects. Many of the methods do not perform well in the presence of multiple structures including ramps, sharp edges, steep slopes and isolated ground points.

Parametric polynomials estimate parameters that best fit the data for a pre-specified family of functions. In many cases, this yields easily interpretable models that explain the underlying structure of the data, but it is not always true. The chosen family of functions can be overly restrictive for some types of data [9]. Fan and Gijbels [10] showed that even a 4th order polynomial fails to give visually satisfying fits. As an alternative, higher order fits may be attempted, but this may lead to numerical instability. The Locally Weighted Regression (LWR) approach can be used as it has many desirable statistical properties including its adaptability to bias problems at boundaries and in regions of high curvature [11]. Fitting within a local neighbourhood considers local point density accurately, which is not always possible for global model polynomial fitting for the whole dataset. Significant point density variation is typical in laser scanner point clouds that can create problems, e.g. when the data resolution decreases, it is harder to separate the ground points [8]. In particular, for steep slopes, this type of global

parametric model fitting may lead to misclassification results and local fitting typically gives better results.

We propose algorithms based on Robust LWR (RLWR) that uses a locally weighted interpolation function based on the local neighbourhood for each and every point. An iterative process is used in which, for each iteration, a predefined robust weight function is used according to the residual values for the fits. Inclusion of a robust weight function in the proposed algorithm makes the estimates robust and down-weights the height errors of the points w.r.t. the fit for the intermediate steps in a robust fashion.

This paper introduces and extends robust variants of LWR methods proposed recently [12]. It is easy to get the lowest points from LiDAR data (lowest z values) but it is challenging to detect outliers that are not part of the ground (whether above or below). This can include points from small objects (horizontal poles), cables, birds etc. along with spurious isolated points. A major advantage is that the proposed methods can extract the ground surface in the presence of near ground outliers as well as those well above or below the ground. The paper also presents comprehensive results with detailed analysis and a variety of datasets consisting mainly of complex urban objects.

We review the relevant literature in Section II. Section III presents related principles and methods used in the proposed algorithms. In Section IV three algorithms are proposed for the classification of ground and non-ground points. The performance of the algorithms is demonstrated and evaluated using artificial data as well as several real Mobile Laser Scanning (MLS) datasets in Section V. Section VI concludes the paper.

II. LITERATURE REVIEW

Existing filtering methods can be categorized into four general groups: (i) morphological filtering, (ii) surface based filtering, (iii) progressive densification, and (iv) segment based filtering. Morphological filtering uses the concept of mathematical morphology. Lindenberger [5] introduced one of the first morphological filtering methods, in which initially, a rough ground surface is extracted by using a seed point based on the assumption that the lowest point or seed belongs to the ground. Then the rough

terrain is refined with an auto-regression process. This algorithm is vulnerable to the size of the structure element [13]. Vosselman [14] developed a slope-based filter incorporating the idea of maximum admissible height difference between two points as a function of the distance between the points. Zakšek and Pfeifer [15] noticed that although the morphologic filtering algorithm is effective in areas with small differences it does not perform as well in areas with steep slopes.

Surface based filtering algorithms start by considering all the points belonging to the ground surface and gradually removes points that do not fit with a general surface model. Kraus and Pfeifer [6] introduced a surface based technique, using robust interpolation and linear prediction, which is an iterative process based on linear least squares interpolation. It determines an individual weight between 0 and 1 for each irregularly distributed point in the dataset in such a way that the modelled surface represents the terrain. Finally, all the data points are classified into ground and non-ground points based on a predefined height difference threshold value. Zakšek and Pfeifer [15] claimed that a robust interpolation method is more efficient than morphologic filtering for steep slopes covered by forest. Akel et al. [16] proposed a method using orthogonal polynomials, pointed: in contrast to other interpolation methods, orthogonal polynomials are not affected by truncation errors, round-off errors, ill-conditioned cases and unstable systems. The use of a high-degree interpolation function makes it possible to fit a global function that can describe the terrain at a given level of detail. Fan and Gijbels [10] claimed higher order fits can lead to numerical instability.

Progressive densification algorithms start with a small subset of the data and iteratively increase the amount of information used to classify the whole dataset. Axelsson [17] introduced a progressive Triangular Irregular Network (TIN). The algorithm uses the lowest point in large grid cells as the seed for the approach. Subsequently, the first subset is triangulated in order to form a reference bare earth surface. Then, for each of the triangles within the TIN an additional terrain point is included if certain criteria are fulfilled. This iterative process continues until no further points can be added to the TIN.

Segmentation and clustering approaches classify whole segments (homogeneous regions) rather than one single point. This approach classifies segments (a group of points) based on local geometrical relations such as height, slope or curvature in a neighbourhood [18]. Tovari and Pfeifer [19] proposed a

two-step segmentation algorithm that starts from a seed point for region growing, examines k neighbour points to see whether they fulfil certain criteria, and then uses robust interpolation for point groups. Pfeifer and Mandlburger [2] pointed out that segmentation based algorithms are better in the built environment (houses, streets, dams, embankments), and are not affected by edge effects for edge based clustering [20].

Kobler et al. [21] introduced the 'Repetitive Interpolation' filter that works on a pre-filtered dataset. Some filtering algorithms in digital image processing use fast neighbourhood operations. However, these can result in loss of precision [22], and may lead to undesired effects. For example gaps can occur caused by occlusion [3]. For an overview of filtering methods see [2−4], [8].

III. RELATED PRINCIPLES AND METHODS FOR THE PROPOSED ALGORITHMS

The new algorithms in this paper use regression analysis, and employ robust and diagnostic regression approaches. This section presents basic ideas of regression, robust, diagnostics and weighted regressions.

Regression analysis is a statistical tool used in many subjects (e.g. computer vision, pattern recognition, photogrammetry and remote sensing) for fitting a model to observed variables [12], [23], [24]. The linear regression model is:

$$Y = X\beta + \epsilon,$$

(1)

where Y is a $n \times 1$ vector of responses, X is a $n \times (m + 1)$ full rank matrix of m explanatory variables including one constant column of 1, β is a $(m + 1) \times 1$ vector of unknown parameters, and ϵ is a $n \times 1$ vector of identically and independently distributed random errors/noise each of which follows a Gaussian normal distribution with zero mean and constant variance σ^2. The well-known Least Squares (LS) method estimates regression parameters by minimizing the sum of the squared residuals, i.e.

$$minimize_{\hat{\beta}} \sum_{i=1}^{n} r_i^2,$$

(2)

where $r_i = y_i - \hat{y}_i$ is the ith residual, and $\hat{y}_i$ is the fitted value of the ith response variable. Optimal

properties have made the LS method the most popular form of regression, but the method produces misleading results if outliers are present in the data [25].

To reduce the influence of outliers, robust regression and regression diagnostics are two complementary approaches. Robust regression fits a regression model to the majority of the data and then finds outliers defined as those points that possess large residuals w.r.t. the robust output, and regression diagnostics is designed to detect and delete or refit (if necessary) the outliers first and then fit the good data using the LS method [25]. Least Median of Squares (LMS) regression, Least Trimmed Squares (LTS) regression, and Weighted Least Squares (WLS) regression are the three most popular robust regression techniques [25]. LMS minimizes the median of the squared residuals instead of minimizing the sum of the squared residuals, i.e.

$$minimize_{\hat{\beta}} \text{median}_i r_i^2.$$

(3)

This estimator effectively ignores almost half of the observations having the largest residuals. LTS regression is defined as:

$$minimize_{\hat{\beta}} \sum_{i=1}^{h} r_i^2,$$

(4)

where $r_1^2 \leq \cdots \leq r_{n/2}^2 \leq \cdots \leq r_h^2 \leq \cdots \leq r_n^2$ are the ordered squared residuals (the residuals are first squared and then ordered). The LTS ignores the largest $(n - h)$ squared residuals in the summation, allowing the fit to ignore $(n - h)$ outliers. Both the methods achieve the highest possible Breakdown Point [25] of 50%, when h is approximately $n/2$. WLS finds outliers first by regression diagnostics [26] and then assigns a weight to each point according to the *outlyingness* of the point as:

$$w_i(x) = \begin{cases} 0, & \text{if the } i^{th} \text{ point is an outlier,} \\ 1, & \text{if the } i^{th} \text{ point is not an outllier.} \end{cases}$$

(5)

Then the model is re-fitted by using LS. Hence, WLS can be defined as:

$$minimize_{\hat{\beta}} \sum_{i=1}^{n} w_i r_i^2.$$

(6)

Among many of the robust and diagnostic regression methods [25], [26], we employ LMS and LTS regression because they are highly robust.

To explore the influence of outliers and to show the advantages of robust regression, we demonstrate LS, LMS and LTS regression on simulated data. In Fig. 1 (a and b), we create two small datasets of 10 points in 2D space. Regular points are in black and red points are outliers. We fit a simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$. Results are shown in Table I and in Fig. 1. In Fig. 1a, the outlier fits the linear pattern but is far from the regular points and is called a good leverage point [25], [26]. This figure shows that all the fitted lines for the different methods are almost in the same direction although they have slightly different parameter values as shown in Table I. In Fig.1b the outlier does not fit with the regular points and we see that LS failed to find the pattern of the majority points. However, the lines produced using LMS and LTS are similar to the LS line without the outlier i.e. are not affected by the outlier. For Fig. 1b, the coefficient of determination R^2 that measures the ability of the fitted model to represent the observed data for the regression method, is only 7.40% for the LS model and increases to 88.84% without the outlier. This is the same for Fig. 1a without the outlier. For Fig. 1a, LS has a larger value of 97.46% for the fit with the outlier than without the outlier that is 88.84% showing that an outlier can lead to inaccurate decisions and it may produce a lower value of R^2 e.g. Fig. 1b. In spite of the presence of different outliers, results for the two datasets are similar for robust (LMS and LTS) regression.

Fig.1.

TABLE I

IV. PROPOSED ALGORITHM

The proposed ground surface points filtering algorithm can be considered as a robust interpolation method within the group of surface based filtering methods [3], [6]. It couples the ideas of LWR and *robustification* of the LWR. It finds the fine detail in the point cloud by smoothing and classifies in-ground (terrain) and non-ground points (off-terrain objects: buildings, trees, walls, poles, etc.).

A. Locally Weighted Regression

LWR termed 'lowess' (Locally WEighted Scatterplot Smoother) or 'loess', is a nonparametric statistical approach introduced by Cleveland [27] and improved by Jacoby [28] and Loader [29]. It models regression functions of independent variable(s) and the dependent variable without any prior

specified functional relationship between the variables. A regression model is determined by fitting parametric functions locally in the space of the independent variables using weighted least squares in a moving fashion [30].

Let y_i and $x_i = (x_{i1},\ x_{i2}, \dots, x_{im})$; $i = 1, 2, \dots, n$ be the measurements of dependent and independent variables respectively. Assume that the dataset is modelled as:

$$y_i = g(x_i) + \epsilon_i,$$

(7)

where ϵ_i are independent and normally distributed with zero mean and variance σ^2, and $g(x_i)$ is a smooth function of x_i. LWR gives an estimate $\hat{g}(x_i)$ at any value of x_i in the space of independent variables. LWR is nonparametric in the sense that it does not specify the functional form of the whole dataset and no specific assumption is made globally for $g(x)$ but locally around a point x_i. We assume that $g(x)$ can be well approximated by a member of a simple class of parametric functions. According to Taylor's theorem, any differentiable function can be approximated locally by a straight line. To estimate $g(x)$ at a point x_i, LWR uses a local neighbourhood $N(x_i)$ of k $(1 \leq k \leq n)$ observations in x space which are closest to x_i. A smoothing parameter $\propto$ $(0 < \propto < 1)$ determines the size of k, which gives the proportion of points that is to be used in each neighbourhood for local regression. A larger local neighbourhood i.e. larger $\propto$ makes the fit smoother, but a smaller local neighbourhood gives a more robust fit. Every point in the local neighbourhood is weighted according to its distance to the interest point x_i. Alternatively, a local neighbourhood can also be considered as a window or fixed distance $h(x)$, and a smoothing window $x_i \pm h(x)$ may be used for fitting a point x_i. If the same number of observations is on either side of the interest point, the weight function is symmetric, otherwise it is asymmetric. A linear or non-linear polynomial function e.g. a quadratic, of the independent variables can be used to fit the model using the Weighted Least Squares (WLS) method. The local parametric function should be chosen to produce an estimate that is sufficiently smooth without distorting the underlying pattern of the data. LWR typically uses a 'tricube' weight function $w(x)$ for the LS fit defined as:

$$w_i(x) = \begin{cases} \left[1 - \left(\dfrac{d(x_i,x_j)}{max_{j \in N(x)} d(x_i,x_j)}\right)^3\right]^3 & ; \quad j \in N(x_i), \\ 0 & ; \quad j \notin N(x_i), \end{cases}$$

(8)

where $d(x_i, x_j)$ is the distance between x_i and x_j in x-space. Fig. 2 shows the 'tricube' weight function. The value of $w_i(x)$ is a maximum for the point closest to x_i and reduces to zero for the kth nearest point x_j to x_i. Points that are too far away with zero weights will be classified as outliers and deemed influential on the analysis. Among the weight functions that decrease to 0, the tricube has been chosen because it enhances a Chi-squared distributional approximation of an estimate of the error variance. Moreover, it provides an adequate smooth result in almost all situations [27].

Fig. 2.

The estimates of the parameters of Eq. (7) are the values of the parameters that minimize:

$$\sum_{i=1}^{n} w_i(x)\,(y_i - g(x_i))^2.$$

(9)

The coefficients from each local neighbourhood are used to estimate the fitted values $\hat{g}(x_i)$ at x_i. The ordered pairs of $x_i, \hat{g}(x_i)$ give the fitted regression line.

B. Robustification of Locally Weighted Regression

LWR may be strongly influenced by outliers because of its LS nature and hence can give inaccurate non-robust results. The problems of outliers are compounded by the fact that the local regressions typically involve a subset of the complete dataset. Therefore, any erroneous data point will compromise a significant proportion of the points used in the local estimation and their degree of influence may cause false estimates [28]. To reduce the effects of outliers and to get a robust fit of the model we use two alternative approaches: (i) assigning a robust weight to each point in the neighbourhood, which is similar to the regression diagnostic approach, and (ii) fitting by using robust regression e.g. LMS and LTS, for each point with its local neighbourhood.

Cleveland [27] used the well-known 'bisquare' weight function to get robust locally weighted regression. The bisquare weight function is defined w.r.t. the residuals of the locally weighted fit as:

$$B(r_i^*) = \begin{cases} \left(1 - r_i^{*2}\right)^2 & for \ |r_i^*| < 1, \\ 0 & for \ |r_i^*| \geq 1, \end{cases}$$

(10)

where $r_i^* = r_i/6$ MAD, MAD is the median of $|r_i|$, and $r_i = y_i - \hat{g}(x_i)$. The shape of the bisquare weight function in Eq. (10) is shown in Fig. 2. It is slightly narrower than the tricube weight function in Eq. (8). The bisquare weight function is used to estimate the new set of Robust LWR (RLWR) coefficients by minimizing the error sum of squares:

$$\sum_{i=1}^{n} B(r_i^*) w_i(x)(y_i - g(x_i))^2.$$

(11)

The estimated robust coefficients are used to obtain a new set of fitted values for $\hat{g}(x_i)$. This robustness step is repeated until the values of the estimated coefficients converge. Cleveland [27] showed, with a large number of experiments on real and artificial datasets, that two iterations are enough to get the final fit. We repeat the robustness step twice. Inclusion of the robust weight function (bisquare) also reduces the influence of outliers on the fits. We call the method Robust Locally Weighted Least Squares (RLWLS).

In the second approach, robust regression is employed to get the robust fit for all the points in the data. We use LMS or LTS regression instead of LS for the neighbourhood of each point in the data. Locally Weighted LMS (LWLMS) and Locally Weighted LTS (LWLTS) robust regression can be performed using:

$$minimize_{\hat{\beta}} \ \text{median}_i \ w_i(x) \ r_i^2,$$

(12)

and

$$minimize_{\hat{\beta}} \ \sum_{i=1}^{h} w_i(x)(r_i^2)_{i:n},$$

(13)

respectively, where $r_1^2 \leq \cdots \leq r_{n/2}^2 \leq \cdots \leq r_h^2 \leq \cdots \leq r_n^2$ are the ordered squared residuals.

C. Implementation

The laser scanning point clouds considered in this paper are acquired along transport corridors using

vehicle mounted laser scanners i.e. MLS. The dataset is typically sliced into manageable 'stripes' for processing and then the results are merged. Ground surfaces such as the road pavements and footpaths are usually considered as the lowest features locally. Belton and Bae [31] defined ground points as the points on the lowest, smooth, nominally horizontal surface. The proposed algorithm finds the lowest level of each point w.r.t. its respective local neighbourhood for every point in a stripe. A characteristic of MLS is the variable point density. We use k Nearest Neighbourhood (kNN) that creates neighbourhoods of constant k because kNN is able to avoid the problem of point density variation. For each stripe of the data, the algorithm processes the two dimensional orthogonal profiles x-z and y-z, which are parallel to and perpendicular to the scanning path for each point respectively. This combination overcomes ambiguities in classification. The algorithm is performed iteratively using two main steps as follows.

First, RLWR is used to get a robust nonlinear fit for the whole stripe. We use linear and not quadratic fitting assuming that, for a sufficiently small size of neighbourhood, linear fitting will be a good approximation to a non-linear or polynomial fit. We can use RLWLS, LWLMS or LWLTS for robust polynomial fitting to the x-z and y-z profiles. The second step consists of the following sequence of four tasks.

Task 1. Calculation of residuals $r_i = z_i - \hat{z}_i$, where $\hat{z}_i$ is the estimate of z_i.

Task 2. Classification of points into two categories: points above the fitted RLWR line and points on or below the fitted RLWR line.

Task 3. The bisquare robust weight function in Eq. (10) is used to down-weight the z-values of the points that are above the fitted line, and the rest of the points are given a weight of one i.e. points on or beneath the fitted line will be unchanged. The reweighted z-values are considered as the new z values for the next fit. Fig. 3 shows the down-weighting process for z_i. If, after using the bisquare weight function, the value of a fitted z is less than the lowest z-value of its neighbourhood, the lowest one replaces the fitted z-value. This is because the lowest features are regarded as the ground surface. However, if the local neighbourhood contains a 'low outlier' [8] then we replace the outlier by the point that has the least z-value among the inlier set. LMS and LTS based LWLMS and LWLTS respectively

can classify the points into inliers and outliers [25] and RLWLS has the ability to ignore the influence of low outliers when it performs *robustification* i.e. uses a bisquare weight function and an iterative process.

Fig. 3.

Task 4. The new set of z-values is used to get the next RLWR fit. Tasks 1 to 3 will be repeated until the difference Δ between the two Root Mean Squared Errors (dRMSE) from the two latest consecutive fitted polynomials is insignificant. That means, threshold of Δ will be less than the distance in z-axis between the two closest points. From a number of MLS data investigations, we see it is sufficient to consider $\Delta= 0.005$ as the threshold value for our algorithms. The final RLWR fit is considered as the lowest or ground level fit for the current stripe and the points between a band created by the lowest level and lowest level $\pm$ a predefined threshold (based on similar data experiments) are considered as ground surface points from the profile. Finally, common points that are identified as ground points from the results of x-z and y-z profiles are classified as the ground points for the stripe. The threshold values for x-z and y-z may vary because the x and y axes measure different directions. For example, in the case of mobile mapping through road corridors, the y axis is the horizontal direction parallel to the road, and the x axis is in the horizontal direction perpendicular to the road. Therefore, the thresholds for a x-z stripe depend on the difference between the points from the two opposite sides of the road and the threshold for a y-z stripe depends on the difference between the points of the two most distant positions on the stripe along the road.

The ground surface extraction workflow is shown in Fig. 4, and the proposed method for x-z profiles detailed in Algorithm 1. The algorithm will be performed in the same fashion for y-z profiles.

Fig. 4.

Algorithm 1.

V. Experiments and Evaluation

We evaluate the RLWR fitting using simulated data, and then demonstrate, compare and evaluate the proposed algorithms for filtering ground surface points on real MLS point cloud data.

A. Robust Locally Weighted Regression Fitting

Fig. 5 shows a simulated 2D dataset of 141 points with one outlying point, generated in a similar way to Moran [32]. A linear functional relation $y = \beta_0 + \beta_1 x$ is used to create a dataset that is non-linear as a whole but locally linear. That means, the values of β_0 and β_1 are fixed for each individual local region but may differ for other regions. The variable x follows a uniform distribution within an interval. To make the dataset noisy, we add an error ϵ variable with the linear model. The noisy data points generated are classified into seven groups following different local linear models:

$$y = \begin{cases} 2.45 + 0.08x + \epsilon & 0 \le x \le 15, \\ 3.1 + 0.035x + \epsilon & 15 \le x \le 35, \\ 6.8 - 0.07x + \epsilon & 35 \le x \le 55, \\ 3 + 0.0x + \epsilon & 55 \le x \le 75, \\ -0.7 + 0.05x + \epsilon & 75 \le x \le 110, \\ 3.3 + 0.01x + \epsilon & 100 \le x \le 120, \\ 5.7 - 0.01x + \epsilon & 120 \le x \le 160, \end{cases}$$

$$(14)$$

where ϵ follows a Gaussian normal distribution with zero mean and standard deviation 0.35. The outlier point is created so that it is far from a local region. We perform LWLS and RLWLS regression. Fig. 5a shows the fitted lines for LWLS (blue dotted) and RLWLS (red) for the whole dataset of 141 points. The RLWLS line is smoother than the LWLS line. Fig. 5b shows that an interest point, which is an outlier (green, filled circle), has been fitted locally with its neighbours (points within the vertical dotted lines). The RLWLS fit is closer than the LWLS fit to most points in the local neighbourhood, which means RLWLS is robust in the presence of outliers.

Fig. 5.

To investigate the necessity for local fitting and to determine the fitting performance of robust methods: RLWLS, LWLMS and LWLTS for a dataset of unspecified and unknown pattern whether it is noisy or noise-free, we generate 140 noise-free data points that are a combination of seven straight lines as shown in Fig. 6a, but does not follow a specific linear model as a whole. This dataset has been made noisy and non-smooth by adding 140 noisy points (red) as shown in Fig. 6c. Fig. 6d shows it is now difficult to observe the underlying seven lines. Figs. 6 (b and d) show that parametric LS linear model does not represent the real line through the points for Figs. 6a and 6c respectively, and is not

representative of the whole dataset. Hence we need to fit the data locally within their respective neighbourhoods. We perform LWLS and RLWLS regression for linear and quadratic functions. Fig. 7a shows the result of locally weighted regression. The LWLS fit does not match in many places. RLWLS using the linear functions gives a smoother line than the RLWLS using quadratic functions. Now we employ LMS and LTS regression for the local regions, and perform LWLMS and LWLTS. Fig. 7b shows that robust regression based methods: LWLMS and LWLTS perform similarly to RLWLS, and closely represent the real line without the noise.

Fig. 6.

Fig. 7.

B. Ground Surface Points Filtering

In this section, the proposed filtering algorithms are demonstrated and evaluated on real MLS datasets collected by a vehicle borne laser scanning system moving at traffic speed in urban areas. We assess the results visually and quantitatively by comparing with those for the robust segmentation algorithm proposed in Nurunnabi et al. [18], and later used in [33], [34] that has been shown to outperform other methods with regard to ground truth. In this case manual inspection of the scanned site and visually examining the raw data and results means these results can be regarded as ground truth when determining the performance. Note that a full survey is not possible because of the need for road closures, interference with normal business activities, and the time taken to accurately survey the area.

To measure the quantitative performance of the proposed filtering algorithms, we follow the rules in Sithole and Vosselman [8] and calculate Type I and Type II errors, total error and accuracy, defined as:

$$\text{Type I error} = \frac{b}{a+b},$$

$$(15)$$

$$\text{Type II error} = \frac{c}{c+d},$$

$$(16)$$

$$\text{Total error} = \frac{b+c}{e},$$

$$(17)$$

$$\text{Accuracy} = \frac{a+d}{e},$$

$$(18)$$

with a: number of ground points correctly identified, b: number of ground points incorrectly identified as non-ground points, c: number of non-ground points incorrectly identified as ground points, d: number of non-ground points correctly identified, and e: total number of data points.

We calculate the measures and compare them with the robust segmentation algorithm [18]. Segmentation groups homogeneous points and labels them as the same region. A region growing approach starts from a seed point that has minimum curvature $\sigma(p)$ [18]. Regions grow based on heuristic and proximity criteria where local normals and curvatures calculated based on the k Nearest Neighbourhood Np_i of every point p_i. The reader is referred to Nurunnabi et al. [18] for details of the segmentation algorithm.

We compare the number of ground g and non-ground ng points extracted from every method.

1) Dataset 1: Pole-tree-wall

This dataset of 32,822 points contains a tree, a light pole, a signpost, part of a road side wall, part of a roof that overlaps the tree and road surfaces, and has sharp edges and steep slopes as shown in Fig. 8a. We use RLWLS, LWLMS and LWLTS on the two bi-dimensional x-z and y-z profiles of the dataset. We use LWR with the tricube weight function for every point in the dataset with the local neighbourhood k of 300 based on knowledge about the data density and from experiments on similar data. We fit the regression line, calculate residuals $r = z - \hat{z}$, and perform the down weighting using the bisquare robust weight function to reduce the influences of extremely high off-terrain points. The iteration process continues until the difference between two Root Mean Squared Errors Δ from two consecutive fits is less than 0.005. Figs. 8(b−g) demonstrate iterative fits for RLWLS, LWLMS and LWLTS based on x-z (Figs. 8b, d and f) and y-z (Figs. 8c, e and g) profiles. The results in Figs. 8 (b and c) show that using RLWLS after six and five iterations respectively we get the ground level for x-z and y-z profiles. In Figs. 8(b and c), the fitted lines from the first, second, third and final iterations are shown in red, yellow, green and magenta respectively. The final iteration (magenta line) derives the

ground level. Figs. 8(d and e) and 8(f and g) show the results of fitting using LWLMS and LWLTS respectively for *x-z* and *y-z* profiles superimposed on the relevant point cloud (in grey). The required iteration numbers for the respective methods are given in Column 12 in Table II. Each successive iteration shows improving accuracy for ground level estimation.

Fig. 8.

After estimating the ground level, we add a distance threshold δ to the *z* value of the ground level. For MLS data, threshold values depend on the width and length of the respective stripes/profiles that may be different. Analysis of similar data is used to decide the values of the threshold. For this dataset, points vertically within 0.30m and 0.35m of the estimated ground level for *x-z* and *y-z* profiles respectively are treated as ground surface points. Fig. 9 shows the results for RLWLS (Fig. 9a), LWLMS (Fig. 9b) and LWLTS (Fig. 9c) of the classified ground (grey) and non-ground points (blue) for *x-z* profile. Identified ground points for the *y-z* profiles can be practically regarded as identical for the methods from a visual inspection. Figs. 10(a, b and c) show the final results for the common ground points from the *x-z* and *y-z* profiles for the three robust methods: RLWLS, LWLMS and LWLTS. Again the results are practically identical from a visual inspection and Table II demonstrates this quantitatively.

Fig.9.

We employ the segmentation algorithm of Nurunnabi et al. [18] and compare the filtering results with the segmentation results. The segmentation algorithm using parameters of neighbourhood size $k = 50$, angle threshold $\theta_{th} = 10^o$, and minimum region size $R_{min} = 2$; produce the result in Fig. 10d. Each region is shown in a different colour and the ground is well defined as three regions: pavement, kerb and road from which we can determine the number of points defined as the ground. We calculate ground *g* and non-ground *ng* points for the different methods and count the points that match with the results from the segmentation algorithm. Table II shows the results are similar for all three robust methods and have error rates of less than 1% and accuracy rates of approximately 97%.

Fig.10.

We calculate the time to perform all the methods using the MATLAB® profile function. Time in

seconds(s) for the respective ground surface extraction methods are given in the last column of Table II. RLWLS takes significantly less time than LWLMS and LWLTS without a reduction in the quality of the results. RLWLS (336.51s) takes significantly less time than the segmentation (724.95s) algorithm that is important if extracting the ground is one of the main objectives. RLWLS is 2.21 and 6.66 times faster than LWLMS (744.88s) and LWLTS (2239.74s) methods respectively.

TABLE II

2) Dataset 2: Traffic signal

Meng et al. [35] pointed out that errors are mainly found in difficult to recognize low height features such as bushes, short walls, and on the boundaries of the ground and non-ground objects. It is also more difficult to identify ground points in an area covered by dense urban features such as power poles, traffic poles/signs and cars. The dataset shown in Fig. 11a consists of 70,272 points for an urban area containing a light pole, cars, cylindrical and planar surfaces, bushes and complex objects including traffic signals, and a smooth slope or ramp.

We run RLWLS, LWLMS and LWLTS for the two *x-z* and *y-z* profiles. We fit LWR, and perform the algorithms by using tricube weight as the local weight, and bisquare weight function for down weighting the extremely high *z* values for every point with neighbourhood size 300. The iteration process terminates at $\Delta < 0.005$. Ground surface points were found below 0.8m and 0.3m from the estimated ground level for the *x-z* and *y-z* profiles respectively. The final ground points (grey) i.e. the common ground points from *x-z* and *y-z* profiles for RLWR, LWLMS and LWLTS are plotted in Fig. 11 (b, c and d) respectively. We also run the segmentation algorithm [18] using parameters: $k = 50$, $\theta_{th} = 5°$, and $R_{min} = 2$, and the results are considered as ground truth as for the pole-tree-wall dataset. Qualitatively, the results can be regarded as comparable from visual inspection and this is backed up by quantitative results in Table III. Table III shows the proposed algorithms have Type-I error, Type-II error and total error rates of less than 0.61%, 2.22% and 1.03% respectively for RLWLS, LWLMS and LWLTS, with more than 97.00% accuracy for ground surface extraction. The time required for RLWLS is again significantly less than for segmentation and other robust methods.

Fig. 11.

TABLE III

3) Dataset 3: Filling-station

We now evaluate the performance of the proposed algorithm in the presence of low or below ground outliers. These are points that normally do not belong to the landscape and have originated from multi-path errors and errors in the laser range finder [8]. The problem with low outliers is that most of the filtering algorithms assume that the lowest points belong to the terrain, and some others assume that points near to a lower point must belong to an object [8], [31]. In the presence of low outliers these assumptions may fail and misclassify ground points into non-ground points. The proposed robust methods: RLWLS, LWLMS and LWLTS can identify low outliers as well as high outliers. Most algorithms are able to deal with high outliers because they are so far above the neighbouring points [8], and our methods easily handle these by down weighting them in consecutive iterations.

We consider 42,032 laser scanning points of a road side filling-station. Fig. 12a shows the dataset has scattered low outliers indicated by the isolated red points as well as one cluster of low outliers. We only consider the RLWLS method here because we have shown that RLWLS is significantly faster than LWLMS and LWLTS for the same performance. We also run LWLS to illustrate outlier influence on the non-robust approach.

The algorithms use the same weight functions as before with a local neighbourhood size k of 500. The iteratively fitted lines for the x-z profile for LWLS and RLWLS are shown in Figs 12b and 12c respectively. Fig. 12b shows that the final fitted line i.e. the estimated ground level (magenta) for LWLS are not free from the effects of low outliers, so the estimated ground level wrongly passes through the low outliers. The result in Fig. 12d, shows that many ground points are misclassified as non-ground (purple) points because of the presence of low outliers. Fig. 12c shows that the final fitted line for the x-z profile from RLWLS successfully ignores the low outliers and estimates ground level accurately. The results in Fig. 12e qualitatively show that non-ground and ground points are properly classified. Hence, the final ground surface extracted with RLWLS is free from non-ground points. We perform the segmentation algorithm with $k = 30$, $\theta_{th} = 10^o$, and $R_{min} = 2$. Segmentation results are in

Figs. 12f. The quantitative results in Table IV show the accuracy rate for RLWLS is 97.84% and for LWLS is 92.69%, which show that RLWLS extracts the ground surface efficiently even in the presence of low outliers in the data.

Fig. 12

TABLE IV

4) Dataset 4: Road-corridor

We apply the proposed robust ground surface extraction algorithm to a large MLS dataset of 1,054,772 points that covers about 56m of road area. The dataset shown in Fig. 13 consists of large trees, buildings, a bus shelter, small walls, fence, signposts, and many complex objects. This dataset also contains some isolated ground points i.e. presence of discontinuities at the left-bottom corner as shown in Fig. 13. In addition, smooth and steep slopes are present in many places. We break the data into ten slices along the *y*-axis. Again we just use RLWLS using the same parameters as for the previous experiments. The final common ground points of the two different profiles *x-z* and *y-z* in Fig. 13 show that RLWLS efficiently classifies ground (grey) and non-ground (purple) surface points in areas covered by dense urban features. The new method handles complex shapes and isolates large buildings as well as small walls and fences from the ground surface. Very low objects such as cars and benches in the bus shelters are isolated efficiently as non-ground objects. The proposed method is successful and can deal with ground points that are isolated from the main ground region i.e. fits this low density small region as well as high density large regions.

Fig. 13.

5) Dataset 5: Aerial laser scanning

This paper mainly focuses on ground surface extraction in MLS data along vehicle corridors. This data mostly has an almost flat terrain. To demonstrate the performance of the proposed algorithm for non-flat terrain we consider an Aerial Laser Scanning (ALS) dataset. Fig. 14a shows data acquired using LiDAR covering 11,432 square metres area in Bunbury, Western Australia, Australia. The point density is 2.10 points/m^2 with a spacing of 0.69m for 23,954 points of mainly ground, and non-ground objects such as big trees, low vegetation and houses. Manual inspection reveals this dataset also has

many low and below ground outliers. We perform RLWLS using the same parameters as for the previous experiments with neighbourhood size 300. The final ground surface points in Fig. 14d were found below 0.8m and 0.5m from the estimated ground level for the *x-z* and *y-z* profiles respectively. Robust fitted lines for *x-z* profiles for a number of iterations are shown in Fig. 14b. Many low outliers are ignored below the final ground level, i.e. the final iteration in magenta shows the ground level is not influenced by low outliers. Results in Fig. 14c reveal that the proposed method efficiently segments ground (grey) from non-ground (purple) surface points. For clarity, in Fig. 14d we remove the non-ground points from Fig. 14c and get the segmented ground points only. Fig. 14d shows RLWLS was successful for extracting ground points from a non-flat terrain.

VI. CONCLUSIONS

In this paper three variants of RLWR based ground surface points filtering methods: RLWLS, LWLMS and LWLTS are proposed. Results show that the LWR based statistically robust approach is good at extracting ground surfaces in urban areas. Most urban features such as complex large buildings, large trees, walls, sign posts, power poles, traffic signals and vehicles are separated from the ground surfaces in spite of the presence of steep slopes, sharp edges and ramps. Although the RLWR approach is an iterative process using local weights, it runs with a very low number of iterations so minimizes the computation time. RLWLS is significantly faster than LWLMS and LWLTS and produces similar results. The ground surface extraction technique means post-processing tasks that only need to operate on non-ground or ground data e.g. tree finding that only needs to operate on part of the data will be faster. An advantage of the algorithm is that it requires only a few parameters. A major advantage is that it can efficiently handle the presence of low or below ground outliers. Quantitative assessment by comparing the results with a state of the art robust segmentation algorithm gives more than 97% correct classification rate of the ground and non-ground surface points with very low error rates. RLWLS is most appropriate to perform filtering because it is the fastest of the robust methods and about 50% faster than the best alternative segmentation method [18]. When considering the determination of the ground points, the majority of non-ground points can be excluded without filtering out points belonging

to small vertical surfaces such as road kerbs.

REFERENCES

[1] K. Kraus, W. Karel, C. Briese, and G. Mandlburger, "Local accuracy measures for digital terrain models," *The Photogramm. Rec.*, vol. 21, no. 116, pp. 342−354, 2006.

[2] N. Pfeifer and G. Mandlburger, "LiDAR data filtering and DTM generation," in *Topographic Laser Ranging and Scanning*, New York, London: CRC Press, Taylor & Francis Group, 2009, pp. 307−334.

[3] C. Briese, "Extraction of digital terrain models," in *Airborne and Terrestrial Laser Scanning*, Scotland, UK: Whittles Publishing/CRC Press, pp. 135−167, 2010.

[4] N. El-Sheimy, C. Valeo, and A. Habib, "*Digital Terrain Modelling: Acquisition, Manipulation, and Its Applications*," USA: Artech House, 2005.

[5] J. Lindenberger, "Laser-Pro-filmessungen zur topographischen gelandeaufnahme", PhD dissertation, Institute for Photogrammetry, Deutsche Geodatische Kommission, Series C, no. 400, Munich, Germany, 1993.

[6] K. Kraus and N. Pfeifer, "Determination of terrain models in wooded areas with airborne laser scanner data," *ISPRS J. of Photogramm. and Remote Sens.*, vol. 53, no. 4, pp. 193−203, 1998.

[7] Y. Zhou, Y. Yu, G. Lu, and S. Du, "Super-segments based classification of 3D urban street scenes," *Int. J. of Advance Robotic Systems*, vol. 9, no. 1, pp. 1−8, 2012.

[8] G. Sithole and G. Vosselman, "Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds," *ISPRS J. of Photogramm. and Remote Sens.*, vol. 59, no. 1, pp. 85−101, 2004.

[9] M. Avery, "Literature review for local polynomial regression," 2012. http://www4.ncsu.edu/mravery/AveryReview2.pdf, Accessed: 20/05/2013.

[10] J. Fan and I. Gijbels, "*Local Polynomial Modelling and Its Applications*," London, UK: Chapman and Hall, 1996.

[11] W. S. Cleveland and C. L. Loader, "Smoothing by local regression: principles and methods," in *Statistical Theory and Computational Aspects of Smoothing*, New York, USA: Springer, 1996, pp.10−49.

[12] A. Nurunnabi, G. West, and D. Belton, "Robust locally weighted regression for ground surface extraction in mobile laser scanning 3D data," *ISPRS Workshop Laser Scanning 2013,* Antalya, Turkey, 11-13 November, 2013.

[13] J. Shan and A. Sampath, "Urban DEM generation from raw LiDAR data: a labelling algorithm and its performance," *Photogramm. Eng. and Remote Sens.*, vol. 71, no. 2, pp. 217−226, 2005.

[14] G. Vosselman, "Slope based filtering of laser altimetry data," in *Proc. Int. Arch. of the Photogramm., Remote Sens. and Spatial Inf. Sci.*, 2000, vol. 33, no. B3, pp. 935−942.

[15] K. Zakšek and N. Pfeifer, "An improved morphological filter for selecting relief points from a LIDAR point cloud in steep areas with dense vegetation," Technical report, Delft Institute of Earth Observation and Space systems, TU Delft, The Netherlands, 2004.

[16] N. A. Akel, S. Filin, and Y. Doytsher, "Orthogonal polynomials supported by regional growing segmentation for the extraction of terrain from LiDAR data," *Photogramm. Eng. and Remote Sens.,* vol. 73, no. 11, pp. 1253−1266, 2007.

[17] P. Axelsson, "DEM generation from laser scanner data using adaptive TIN models," in *Proc. of The Int. Arch. of Photogramm. Remote Sens. and Spatial Inf. Sci.*, 2000, vol. 33, no. B4/1, pp 110−117.

[18] A. Nurunnabi, D. Belton, and G. West, "Robust segmentation in laser scanning 3D point cloud data," *in Proc. of Digit. Image Computing: Tech. and Appl. (DICTA)*, Fremantle, Australia, 3−5 December, 2012.

[19] D. Tóvári and N. Pfeifer, "Segmentation based robust interpolation: a new approach to laser data filtering," in *Proc. Int. Arch. of the Photogramm., Remote Sens. and Spatial Inf. Sci.*, 2005, vol. 36, no. Part3, pp. 79−84.

[20] M. Brovelli, M. Cannata, and U. Longoni, "LiDAR data filtering and DTM interpolation within GRASS," *Trans. in GIS*, vol. 8, no. 2, pp. 155−174, 2004.

[21] A. Kobler, N. Pfeifer, P. Ogrinc, L. Todorovski, K. Ostir, and S. Dzeroski, "Repetitive interpolation: a robust algorithm for DTM generation from aerial laser scanner data in forested terrain," *Remote Sens.of Environment*, vol. 108, no. 1, pp. 9−23, 2007.

[22] P. Axelsson, "Processing of laser scanner data-algorithms and applications," *ISPRS J. of Photogramm. and Remote Sens.*, vol. 54, no. 2−3, pp. 138 −147, 1999.

[23] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, "Robust regression methods for computer vision: a review," *Int. J. of Computer Vision*, vol. 6, no. 1, pp. 59−70, 1991.

[24] C. Bishop, *"Pattern Recognition and Machine Learning"*, New York, USA: Springer, 2006.

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

[26] A. Nurunnabi, A. S. Hadi, and A. H. M. R. Imon, "Procedures for the identification of multiple influential observations in linear regression," *J. of Applied Statistics*, vol. 41, no. 6, pp. 1315−1331, 2014.

[27] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. of the Amer. Stat. Asso.*, vol. 4, no. 368, pp. 829−836, 1979.

[28] W. Jacoby, "Loess: a nonparametric, graphical tool for depicting relationships between variables," *Electoral Studies*, vol. 19, no. 4, pp. 577−613, 2000.

[29] C. Loader, "Smoothing: local regression techniques," in Handbook of Computational Statistics, New York, USA: Springer, pp. 571−596, 2012.

[30] W. S. Cleveland and E. Grosse, "Computational methods for local regression," *Stat. and comp.*, vol. 1, no. 1, pp. 47−62, 1991.

[31] D. Belton and K.-H. Bae, "Automatic post-processing of terrestrial laser scanning point clouds for road feature surveys," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2010, vol. 38, no. 5, pp.74−79.

[32] G. W. Moran, "Locally Weighted Regression Scatter Plot Smoothing (LOWESS): A Graphical Exploratory Data Analysis Technique," Master's thesis, Dept. of Spatial Sci., Naval Postgraduate School, Monterey, California, 1984.

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

[34] A. Nurunnabi, G. West, and D. Belton, "Outlier detection and robust normal-curvature estimation in mobile laser scanning 3D point cloud data," *Patt. Recog.*, vol. 48, pp. 1400−1415.

[35] X. Meng, N. Currit, and K. Zhao, "Ground filtering algorithms for airborne LiDAR data: a review of critical issues," *Remote Sens.*, vol. 2, no. 3, pp. 833−860, 2010.
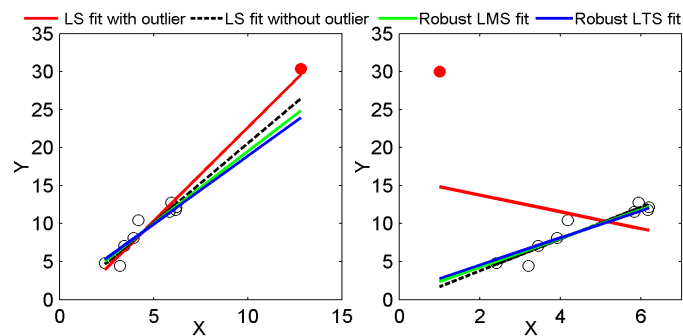
Fig.1. Regression model fitting: (a) outlier in both *x* and *y* directions, and (b) outlier in *y* direction.
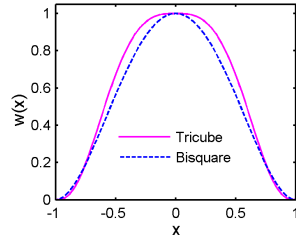
Fig. 2. Tricube and bisquare weight functions.

$$w_i(z) = \begin{cases} \left(1 - r_i^{*2}\right)^2, & for \ |r_i^*| < 1 \\ 0 & , for \ |r_i^*| \geq 1 \end{cases}$$
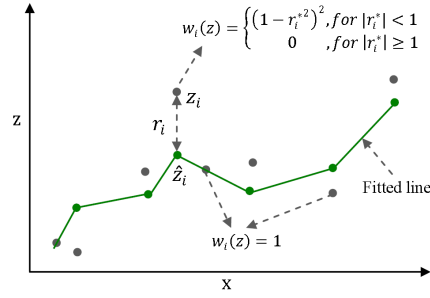
Fig. 3. Down weighting for the *x-z* stripe to get the *z* values for the next fit. Black dots are real points, green dots are the respective fitted points and the green line represents the fitted line.
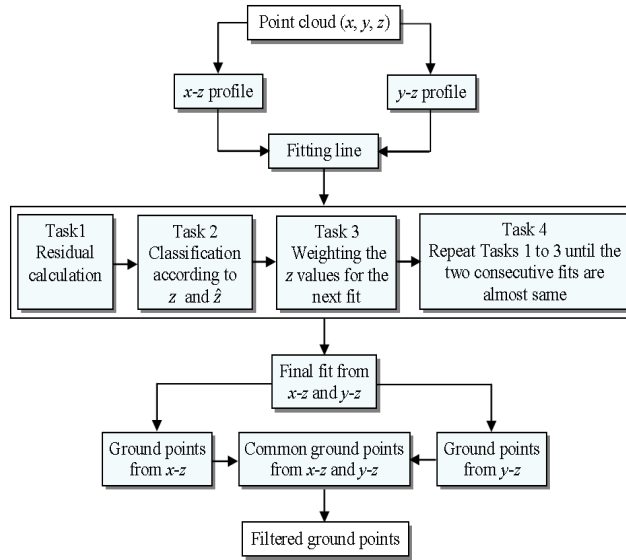

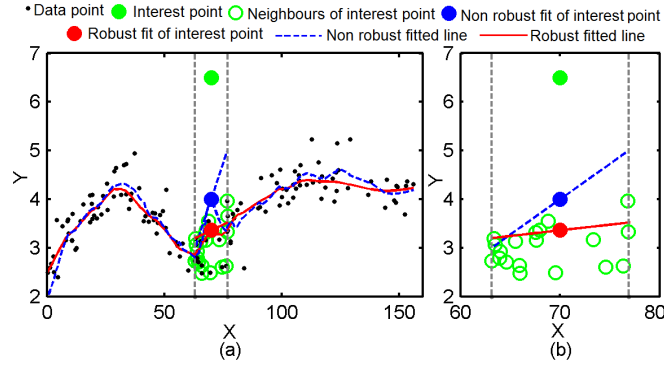Fig. 4. Robust ground surface extraction process.

Fig. 5. Locally weighted regression: (a) fitting for the whole dataset, and (b) fitting only for the interest/outlier point (green, filled circle) with its local neighbourhood.
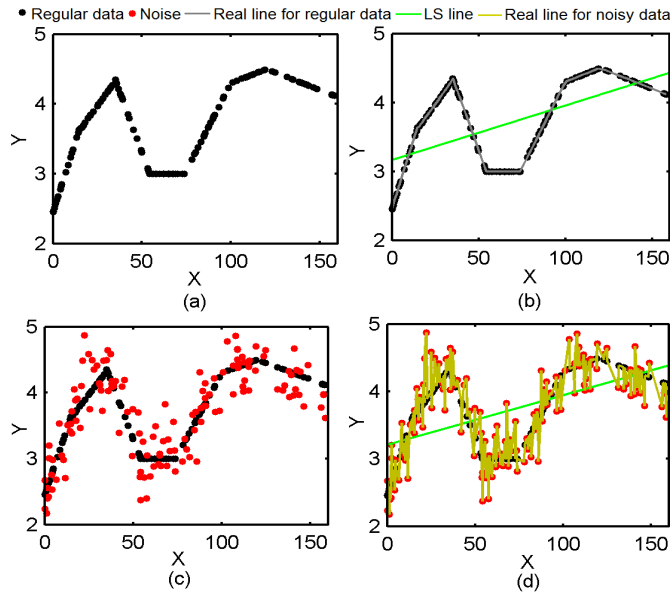

Fig. 6. (a) Simulated regular data of 140 points, (b) real line (black) and LS fitted line (green) for regular data, (c) regular data with 140 noise points (red) added, and (d) real line (olive) and fitted line (green) for regular data with 140 noise points added.
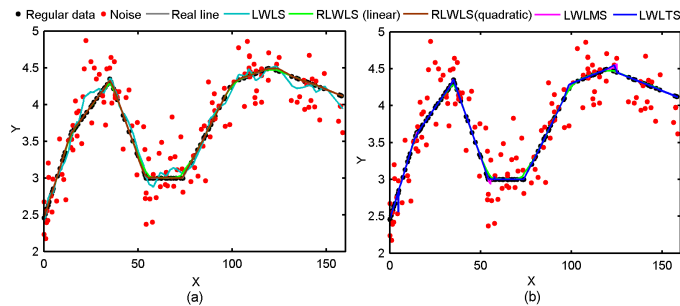

Fig. 7. Simulated data of 140 points with 140 noise points added: (a) real line without noise, and fitted lines for LWLS (blue), RLWLS (linear, green) and RLWLS (quadratic, maroon), and (b) real line without noise, and fitted lines: RLWLS (linear), LWLMS and LWLTS for the data with noise.
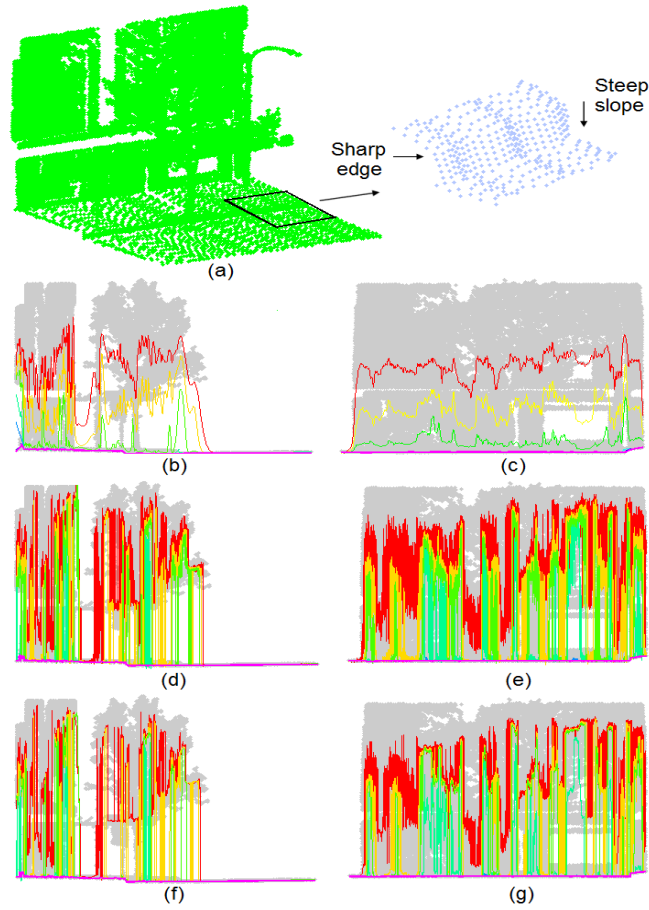
Fig. 8. (a) Pole-tree-wall dataset, iterative fittings for x-z and y-z profiles: (b) RLWLS; x-z, (c) RLWLS; y-z, (d) LWLMS; x-z, (e) LWLMS; y-z, (f) LWLTS; x-z, and (g) LWLTS; y-z.
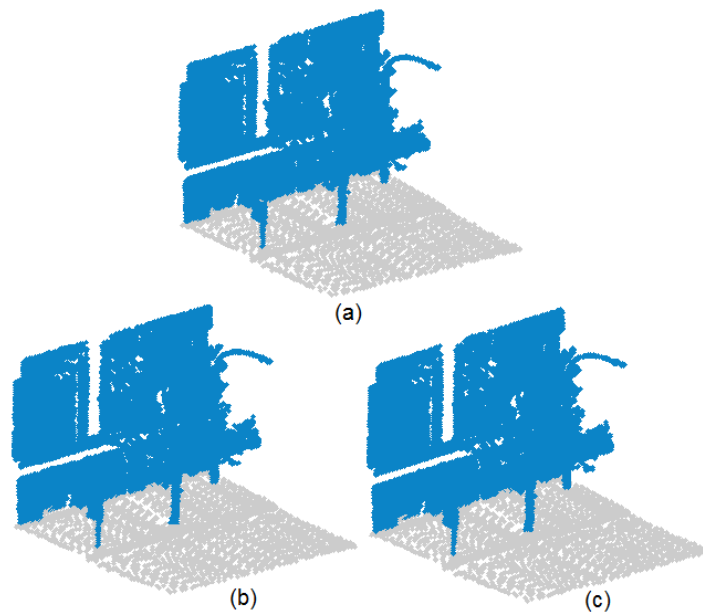

Fig.9. Ground surface extraction for the pole-tree-wall dataset from x-z profiles: (a) RLWLS, (b) LWLMS, and (c) LWLTS.
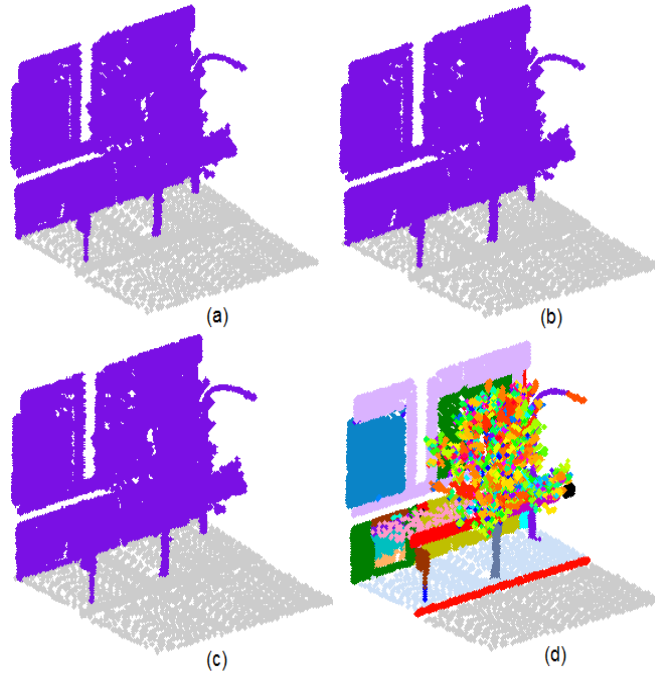
Fig.10. Ground surface extraction for the pole-tree-wall dataset: (a) RLWLS, (b) LWLMS, (c) LWLTS, and (d) segmentation result.
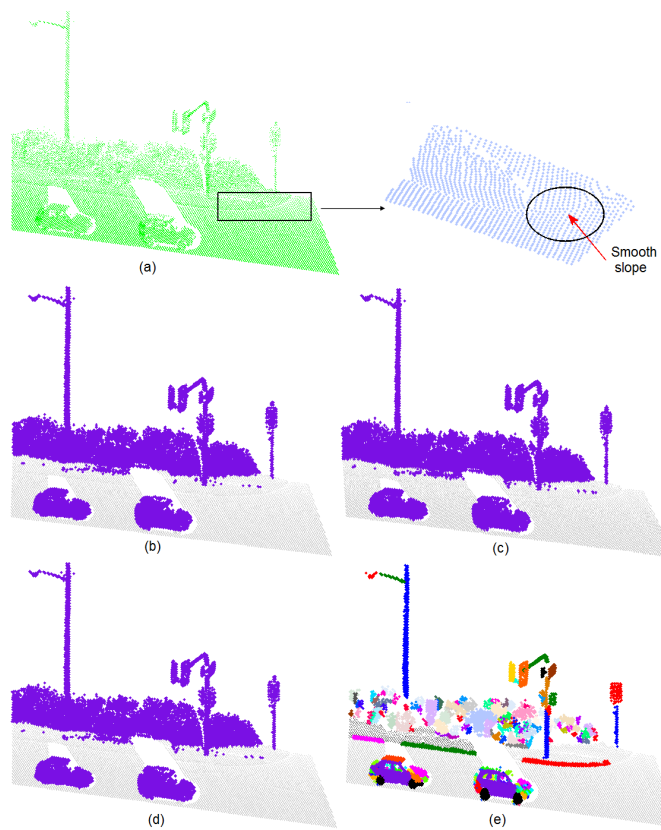

Fig. 11. Ground surface extraction for the traffic signal dataset: (a) dataset. Ground surface extraction results: (b) RLWLS, (c) LWLMS, (d) LWLTS; and (e) segmentation result.

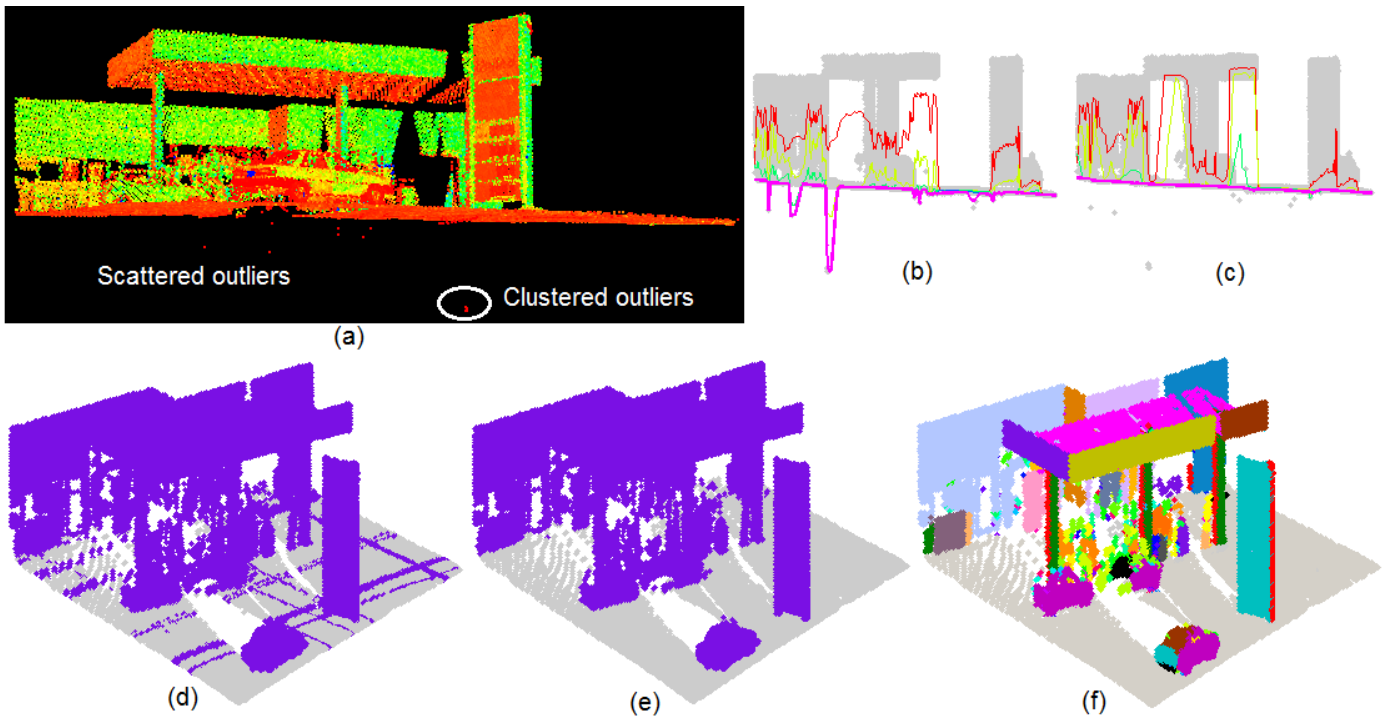**Fig. 12.** Ground surface extraction for the filling-station dataset in the presence of low outliers: (a) real data with low outliers (red points), (b) iterative fittings using LWLS for *x-z* profile, (c) iterative fittings using RLWLS for *x-z* profile, (d) results for LWLS, (e) results for RLWLS, and (f) segmentation results.
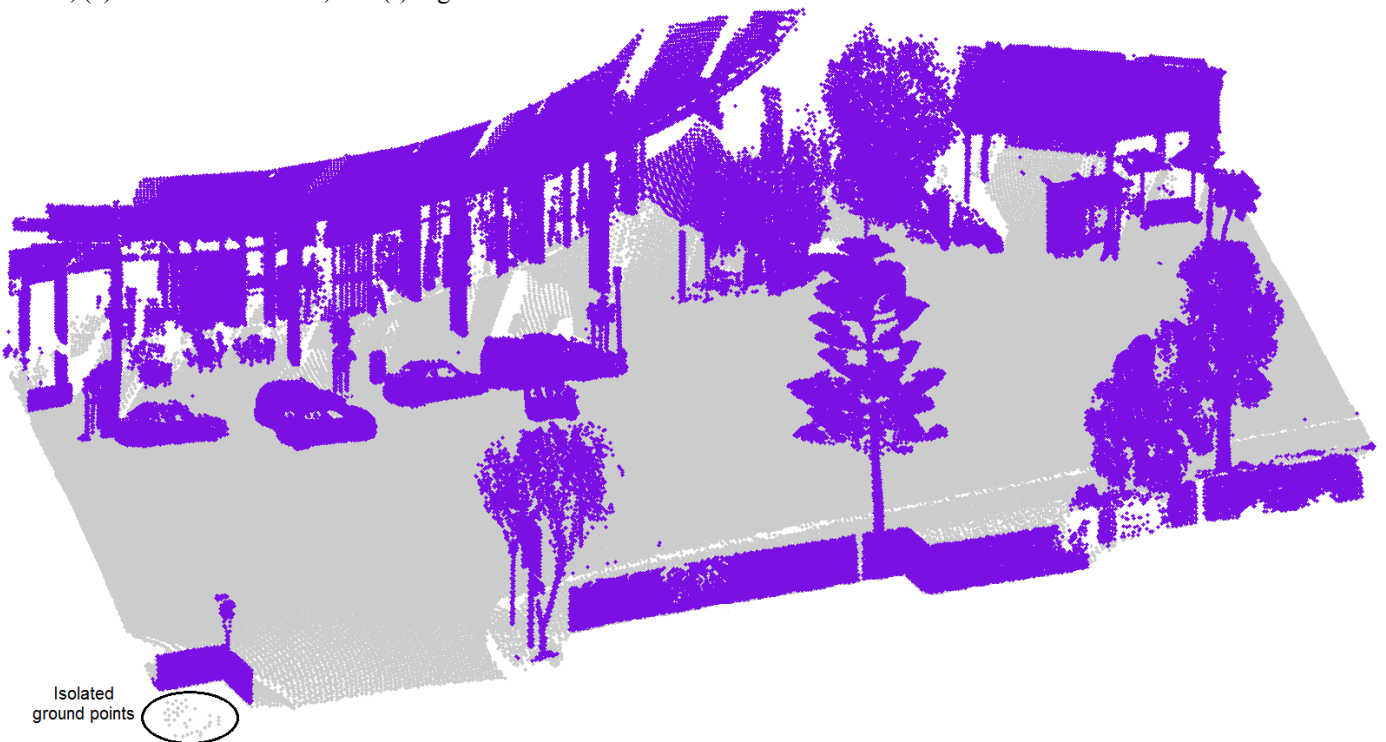

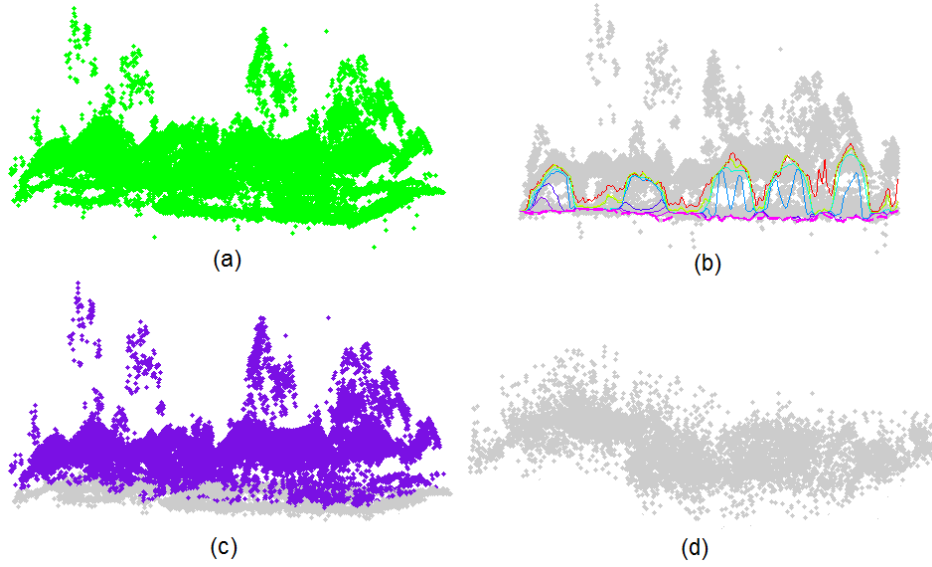Fig. 13. Ground surface extraction for the road-corridor dataset.

Fig. 14. Ground surface extraction for the ALS dataset: (a) dataset, (b) iterative RLWLS fitted lines for *x-z* profile, (c) classification of ground and non-ground points (purple), and (d) extracted ground points.

Algorithm 1: Ground Surface Points Filtering
Input:
P: Point cloud $P(x, y, z)$.
k: Neighbourhood size
Δ: Threshold for difference between two consecutive RMSE
δ: Threshold to get ground surface points
Output:
g: Ground surface points
ng: Non-ground surface points
1.   **for** $i = 1$ to size(P) ***do***
2.   Find k-nearest neighbourhood Np_i for p_i in x-direction
3.   Fit locally weighted regression in the Np_i
4.   **end for**
5.   Find residuals $r_i \leftarrow z_i - \hat{z}_i$
6.   Calculate bisquare weight,w_i using Eq. (11)
7.   **if** z_i larger than $\hat{z}_i$ **then**
8. $z_i \leftarrow z_i * w_i$
9.   **end if**
10.   **if** weighted $z_i < min\ (z_i)$ of Np_i **then**
11. $z_i \leftarrow \min(z_i)$ of (Np_i)
12.   **end if**
13.   Repeat step 3 to step 12 until $|\mathrm{RMSE}_i - \mathrm{RMSE}_{i-1}| < \Delta$
14.   **if** $z_i \leq \hat{z} + \delta$ **then**
15. $g \leftarrow p_i$

16.	**else**
17.	$ng \leftarrow p_i$
18.	**end if**

TABLE I

LS, LMS AND LTS REGRESSION BASED MODEL PARAMETER ESTIMATION.

Dataset	Methods	Coefficients		$R^2(\%)$
		β_0	β_1	
Fig. 1a	LS with outlier	-1.965	2.461	97.46
	LS without outlier	-0.339	2.088	88.84
	LMS	0.498	1.899	99.38
	LTS	1.004	1.788	NA
Fig. 1b	LS with outlier	15.983	-1.105	07.40
	LS without outlier	-0.339	2.0876	88.84
	LMS	0.498	1.899	99.37
	LTS	1.004	1.788	NA

TABLE II

FILTERING ACCURACY MEASURES FOR THE POLE-TREE-WALL DATASET.

Methods	g	ng	a	b	c	d	Type I error (%)	Type II error (%)	Total error (%)	Accuracy (%)	No. of iteration (x-z, y-z)	Time(s)
Segmentation	2886	29205										724.95
RLWLS	2923	29899	2884	2	97	29108	0.0693	0.332	0.302	97.471	6,5	336.51
LWLMS	2928	29894	2884	2	113	29092	0.0693	0.387	0.350	97.422	6,6	744.88
LWLTS	2928	29894	2884	2	113	29092	0.0693	0.387	0.350	97.422	6,6	2239.74

TABLE III

FILTERING ACCURACY MEASURES FOR THE TRAFFIC SIGNAL DATASET.

Methods	g	ng	a	b	c	d	Type I Error (%)	Type II error (%)	Total error (%)	Accuracy	No. of iteration	Time (s)

	g	ng	a	b	c	d	Type I Error (%)	Type II error (%)	Total error (%)	Accuracy (%)	No. of iteration (x-z, y-z)	
Segmentation	49972	19797										1744.55
RLWLS	50222	20050	49123	269	440	19357	0.5446	2.223	1.01	97.450	6, 5	1288.30
LWLMS	50148	20124	49089	303	421	19376	0.6135	2.127	1.03	97.429	6, 6	1755.97
LWLTS	50148	20124	49090	302	421	19376	0.6114	2.127	1.03	97.430	6, 6	5003.37

TABLE IV

FILTERING ACCURACY MEASURES FOR THE FILLING-STATION DATASET WITH LOW OUTLIERS.

Methods	g	ng	a	b	c	d	Type I Error (%)	Type II error (%)	Total error (%)	Accuracy (%)	No. of iteration (x-z, y-z)
Segmentation	11839	29469									
LWLS	10230	31800	9779	2060	286	29183	17.40	0.971	5.581	92.69	5, 6
RLWLS	12359	29661	11834	5	179	29290	0.04	0.607	0.438	97.84	5, 6