Robust Segmentation in Laser Scanning 3D Noisy Point Cloud Data

X
Department of Spatial Sciences
Curtin University, Perth, Australia
CRC for Spatial Information
Email

Y
Department of Spatial Sciences
Curtin University, Perth, Australia
CRC for Spatial Information
Email

Z
Department of Spatial Sciences
Curtin University, Perth, Australia
CRC for Spatial Information
Email

Abstract— **Segmentation is a most important intermediate step in point cloud data processing and understanding. Covariance statistics based local saliency features from Principal Component Analysis (PCA) are frequently used for point cloud segmentation. However it is well known that PCA is sensitive to outliers. Hence segmentation results can be erroneous and unreliable. The problems of surface segmentation in laser scanning point cloud data are investigated in this paper. We propose a region growing based statistically robust segmentation algorithm that uses a recently introduced fast Minimum Covariance Determinant (MCD) based robust PCA approach. Experiments for several real laser scanning datasets show that PCA gives unreliable and non-robust results whereas the proposed robust PCA based method gives more accurate and robust results for planar and non planar smooth surface segmentation.**

Keywords: Covariance technique; feature extraction; outlier; region growing; robust normal; robust statistics

I. INTRODUCTION

Segmentation plays a key role in surface reconstruction, feature extraction, 3D modeling and object recognition. All are strongly interrelated tasks in many areas including pattern recognition, computer vision, image processing, reverse engineering, photogrammetry and remote sensing [Besl and Jain, 1988; Klasing 2009; Liang et al. 2011; Liu and Xiong, 2008; Pauly et al. 2002]. It is the process of separating and labeling the most similar surface points into a number of separate surfaces. Segmentation in point cloud data is not a trivial task because the 3D points in a point cloud are usually unorganized, incomplete, without any connection information, sparse, and have uneven distribution. In addition the surface shape can be arbitrary with sharp features and there is no statistical distributional pattern in the data [Liang et al. 2011; Sotoodeh, 2006]. Moreover, the physical limitations of the sensors, boundaries between 3D features, occlusions, multiple reflectance and noise can produce off-surface points that appear to be outliers [Sotoodeh, 2006]. The true data combined with the presence of outliers make the segmentation process more complex and harder.

Conventional approaches to the segmentation problem can be classified mainly into three categories: edge/border based [Fan et al. 87], region based [Besl and Jain, 88] and hybrid [Woo et al. 2002]. Edge based methods usually detect edge and boundary points, then group the points that are within the identified boundaries and connected edges. Region based approaches use local surface neighborhoods to combine nearby points that have similar properties (orientation etc.) to obtain homogeneity within regions and consequently find dissimilarity between the different regions. Hybrid methods involve both the edge and boundary based approaches to overcome the limitations in both to give better segmentation. It is true that the success of hybrid methods depends on the success of either or or both underlying methods. Scanline [Jiang et al. 1996] and geometric primitive based methods [Marshall et al. 2001] have been introduced for point cloud segmentation. The scanline method is based on grouping the scan lines but has limitations [Powel et al. 1998]. The method is not good for unordered and uneven density point clouds, situations that commonly occur for real data. Klasing [2009] identifies limitations of the geometric primitives based method including the problem of predicting the segmentation results and the high computational cost for large numbers of features. Region based methods are more robust to noise than edge-based methods when using global information [Liu and Xiong, 2008]. In the same paper the authors mention the problem of over or under segmentation, the problem of determining region borders accurately, and the sensitivity of the location of initial seed regions. Similar comments have been made concerning the location of seed points [Chen and Stamos, 2007]. Inaccurate estimates of the normals of range points near region boundaries can cause inaccurate segmentation results. We consider all the points raised in Liu and Xiong [2008] and Chen and Stamos [2007] for finding a reliable segmentation algorithm.

Principal Component Analysis (PCA) based local surface saliency features (e.g. normal and curvature) are used frequently for region growing and edge based methods in segmentation [Pauly et al. 2002; Hoppe et al. 1992]. Since classical PCA uses non-robust location and scatter in its analysis, the consequent estimates (saliency features, e.g. normal and curvature) are erroneous and results in segmentation are inaccurate, unreliable and non-robust. Mitra and Nguyen [2003] state that the sensitivity of PCA to outliers means it fails to accurately fit planes. The literature reveals most of the limitations in the region growing process are related directly and/or indirectly to outlier problems. One of the

1

major problems of outliers is it is not possible to get accurate and robust normals in presence of outliers. If correct normals are robustly estimated for each point, the geometry of even strongly corrupted point-clouds can be perceived [Li et al. 2010]. Robust PCA can provide robust covariance statistics and as a result we get robust normals and curvatures.

To reduce outlier effects on the estimates, this paper proposes a statistically robust segmentation algorithm based on the robust PCA approach. The proposed method is beneficial because of its high resistance to outliers. It can efficiently handle over and/or under segmentation, select robust seed points, detect border/edge points and gaps and reduce outlier effects for the whole process. The algorithm is efficient for non planar smooth surfaces as well as for planar surfaces.

The rest of the paper is organized as follows. Section II briefly discusses the relevant principles and methods. We propose the robust segmentation algorithm in Section III. In Section IV, experiments are performed using real laser scanner data to show the efficiency of the new algorithm followed by conclusions in Section V.

II. RELATED PRINCIPLES AND METHODS

A. Principal Component Analysis

PCA is one of the most popular statistical techniques for dimension reduction and visualization of the data structure. It finds Principal Components (PCs) that are linear combinations of the original variables ranked in terms of the variability in the data given by the variances. The corresponding orthogonal directions re given by the eigenvectors of the Covariance matrix (C) of the data. As a result, the PCs describe the data variances in different orthogonal directions. PCA works as a basis transformation to diagonalize an estimate of the covariance matrix:

$$C_{3\times 3} = \frac{1}{k}\sum_{i=1}^{k}(p_i - \bar{p})(p_i - \bar{p})^T, \quad \bar{p} = \frac{1}{k}\sum_{i=1}^{k}p_i. \qquad (1)$$

In 3D point cloud data, a point $p_i = (x_i, y_i, z_i)$ ($p_i \in P \in \mathrm{R}^3$; P is a dataset), and $\bar{p}$ is the centre of k points. To perform the transformation one has to solve the eigenvalue equation:

$$\lambda V = CV. \qquad (2)$$

Solving the equation by Singular Value Decomposition (SVD) on C, we can easily get the eigenvalues λ ($\lambda_2 \geq \lambda_1 \geq \lambda_0$) and eigenvectors V (v_2, v_1 and v_0). Since C is a symmetric and positive semi-definite matrix all $\lambda_i \geq 0$, and λ_i describes the variation along v_i.

B. Saliency Features

Saliency features are based on the covariance statistics (λ and V in Eq. (2)) PCA has been used for points classification and region growing in segmentation [Pauly et al. 2002; Hoppe et al., 1992]. C in Eq. (1) for a local neighborhood N_{pi} of a point p_i has k neighbours, with the neighborhood size chosen to be sufficiently small to approximate the points as lying in a local planar surface. Since every PC describes as much of the data variance as it can with the highest PC describing the most

variance, the first two PCs are enough to explain a planar surface. Thus v_0 approximates the surface normal n for p_i, λ_0 describes the variation along the surface normal, and the surface variation (curvature) is defined as [Pauly et al. 2002]:

$$\sigma(p_i) = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}. \qquad (3)$$

Most of the existing techniques [Pauly et al. 2002; Hoppe et al. 1992; Pauly et al. 2003; Rabbani et al. 2006; Oehler et al. 2011] use PCA to get the saliency features. Unfortunately the PCA estimates for saliency features are very sensitive to outliers and gives inaccurate and non-robust segmentation results.

C. Robust Principal Component Analysis

To reduce outlier effects and to get robust estimates of saliency features, we use Robust PCA (RPCA) and compute the eigenvalues and eigenvectors in a robust way. Many robust centre (mean) and covariance estimators have been introduced in the literature [Maronna and Yohai, 1998] to get robust variants of PCA. We use the robust PCA proposed by Hubert and Rousseeuw [2005] that combines the idea of Projection Pursuit (PP) with the fast-MCD (FMCD). The PP is used to pre-process the data so that the transformed data are lying in a subspace whose with dimension is less than the total number of data points. T, and then the FMCD [Rousseeuw and Driessen, 1999] estimators are used to get the robust centre and covariance matrix. In PP, if a point is a multivariate outlier, then there must be some one-dimensional projection of the data for which the point is a univariate outlier [Maronna and Yohai, 1995]. Since, "PP uses trimmed global measures; it has the advantage of robustness against outliers" [Friedman nd Tukey, 1974]. The FMCD has better statistical efficiency and it is asymptotically normal [Rousseeuw and Driessen, 1999]. In RPCA, first the data is compressed to the PCs defining potential directions. Then, each ith direction is scored by its corresponding value of "outlyingness". The Stahel-Donoho [Stahel, 1981; Donoho, 1982] measure of outlyingness is defined as:

$$w_i = \arg\max_{\|v\|=1} \frac{\left|p_i v^T - c_{MCD}(p_i v^T)\right|}{\Sigma_{MCD}(p_i v^T)}, \qquad (4)$$

where the maximum is used for all directions, v is a univariate direction and $p_i v^T$ denotes a projection of the ith observation p_i on the direction v. For every direction a re-weighted robust centre c_{MCD} and scatter Σ_{MCD} of the projected data points $p_i v^T$ are computed. Next, a fraction h ($>k/2$) of observations with the smallest values of w_i are used to construct a re-weighted robust covariance matrix Σ. In our experiments, we use $h=0.75k$. Finally, robust PCA projects the data points onto the r-dimensional subspace spanned by the r ($r = 2$ for plane fitting) largest eigenvectors (PCs) of Σ and computes their centre and scatter by the re-weighted FMCD. The eigenvectors of Σ then determine the robust PCs, and the eigenvalues can be used to get the robust saliency features.

In the PCA context, an outlier can be defined in two ways. Orthogonal outlier, that is determined from the orthogonal

distance (the distance between an observation and its projection in the r dimensional PCA subspace) defined as:

$$OD_i = \| p_i - \widehat{p}_i \|, \qquad (5)$$

where $\widehat{p}_i$ is the projection of p_i onto the PCA subspace. The other type of outlier is measured by the scored distance within the PCA subspace. These are then the points that lie far from the bulk of the points within the projected space (e.g. in case of plane fitting, the best fit plane). RPCA has the advantage of being able to find these two types of outliers as a by-product in a diagnostic way.

III. SEGMENTATION

It is known that point cloud data do not follow any statistical distribution, so it is not possible to determine any pattern or mathematical model *a priori*. We formulate the problems in segmentation for identifying underlying patterns in an unsupervised non-parametric fashion.

It is proposed that every surface point in a sufficiently-small local region/surface (neighborhood) can be considered to be lying on a planar surface. This principle allows us to use the local saliency feature information to check the behaviour of a point with respect to a smooth surface. It can be shown that points on some local planar surface may lie on a smooth co-surface under certain conditions. Fig. 1 illustrates the various aspects when considering the segmentation of data into three planes (shown in 1D for clarity).
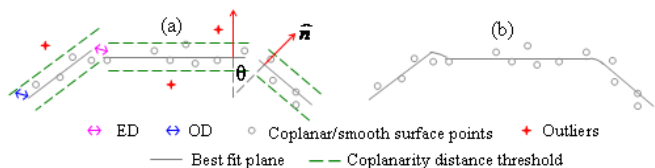

Figure 1.   (a) Planar surfaces (b) non-planar smooth surface

If we think that the gaps between the two boundary points of different planar surfaces are not enough to consider them separate then they may be co-surface points with certain coherence criteria, otherwise discontinuity appears in the gaps. In Fig. 1 (a), we see three planar surfaces that appear to have disjoint/discontinuities (gaps) between them. The first two planes appear to have a jump edge (as in Fig. 2 (b)) and the last two planes appear to have a crease edge (as in Fig. 2 (a)) at their boundaries. In Fig. 1, points near gaps and edges are the major concern when determining which points belong to which surface. Therefore careful analysis is needed around edges and gaps for good segmentation.

A. Problems: Edges, Gaps and Outliers

Edges, gaps and outliers are considered as the three main and unavoidable problems when segmenting point cloud data. Hence a proper understanding about them is vital for providing better solutions for accurate and reliable surface segmentation and reconstruction.

1) Edges: Many authors deal with edges as a means of delineating object boundaries. Usually, three types of edges are considered (see Fig. 2): (a) crease/corner edges such as those where two sides of a roof meet, and where the surface point normals are influenced by different planes, (b) jump/step edges that occur where a surface undergoes discontinuty and close to the boundary points on the two planes have the same orientation, and (c) smooth/virtual edges that can be charaterised by similar normals but there is a change of curvature.
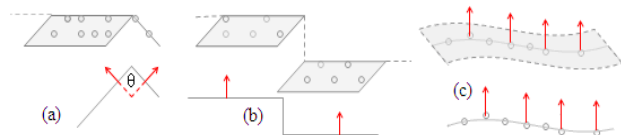

Figure 2.   (a) crease edge (b) jump edge (c) smooth edge

2) Gaps: Unexpected interruption in data acquisition, a faulty sensor, any obstacles hat obstruct the laser pulse and/or uneven surface point density may cause gaps (discontinuties and holes) in the acquired data points. Gaps of the types shown in Fig. 3 are usually seen. Misleading gaps can be filled by boundary extension, which is possible given a thorough analysis of the neighbouring surface points.
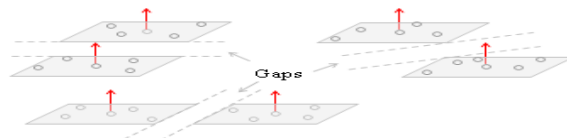

Figure 3.   Gaps in different surface positions

3) Outliers/noise: Outlier issues further exacerbate the problem of edges and gaps. A common idea about outliers is that they are classified as points that are far from the majority points in a dataset, or do not follow the same pattern as the majority follows. Hence in a general sense noise can appear as outliers in many cases. Outliers contaminate local neighborhood based covariance statistics giving inaccurate normals and curvature values. Most of the time, points in a local neighborhood with outliers result in the tangent plane being biased in the direction of the outlier. The presence of outliers in different positions (especially on the edges and boundaries) on a surface results in errors in the estimates of local saliencies (e.g. normals). Fig 4 shows outliers between pairs of surfaces that can join them erroneously, and can change the orientations of the two planar surfaces. Even outliers between two points in a neighborhood can produce erroenous discontinuties in a surface.
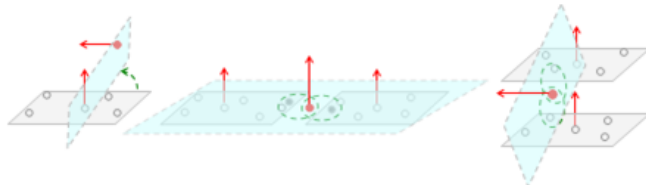

Figure 4.   (a) Outlier as an off-surface point (b) outlier between two co-planar planes (c) outlier between two non-coplanar parallel planes

3

B. Proximity and Coherence Criteria

1) Normals: Visual surface orientation is a useful criterion for feature surface reconstruction and object recognition. This orientation can be represented by the unit normal to the best fitted plane at a point of interest p_i. Surface reconstruction and segmentation depend on normal estimation because the fastest and more robust surface reconstruction algorithms requires points with normals [Yoon et al. 2007]. Many approaches are available in the literature [e.g. Hoffman and Jain, 1987; Hoppe et al., 1992; Wang et al., 2001; Yang and Feng, 2005] for normal estimation. A recent survey [Klasing et al. 2009] shows that the PCA based approach in [Hoppe et al. 1992] is one of the most efficient and popular. To show that PCA is sensitive to outliers and gives inaccurate normals in the presence of outliers/noise, we take a small sample from a planar surface in a point cloud and contaminate it with 20% outliers (Fig. 5(a)). Fig. 5(b) shows the influence of outliers on PCA normals producing errors where outliers are present. In surface reconstruction, the quality of the approximation of the output surface depends on how well the estimated normals approximate the true normals of the sampled surface [Tamal et al. 2005]. Robust and accurate normals are essential to detect and reconstruct sharp/edge features. Yoon et al. [2007] show that outlier normal error estimation is the most likely source of problems when using state-of-the-art surface reconstruction techniques. We use RPCA to reduce the outlier problem. Fig. 5 (c) shows RPCA normals that can be seen to be not influenced by the outliers as they all pointing in a similar direction.
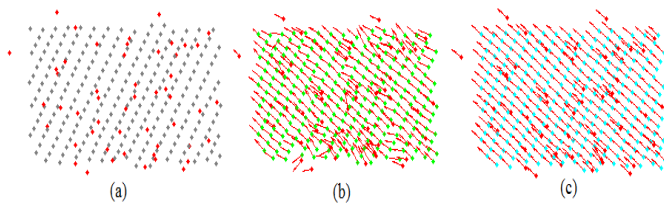

Figure 5.   (a) Point cloud planar surface with outliers (red points) (b) PCA normals (c) RPCA normals

2) Neighborhood: A problem related to proper normal estimation is obtaining the proper neighborhood of a point. Two common neighborhood selection methods are Fixed Distance Neighbors (FDN) and *k* Nearest Neighbors (kNN). kNN has the advantage that it can deal with data that has an uneven point density and can adapt the Area of Interest (AOI) with respect to the data density. We use the K-D tree search algorithm based kNN to get a local neighborhood NP_i of p_i, mainly because of our interest in mobile mapping data (acquired from scanning lasers on vehicles), which has varying point density because of the movement of the vehicle. Neighborhood size is a major concern for reliable local saliency estimation. Hoffman and Jain [1987] note that smaller neighborhoods give normals more susceptible to noise. Many authors suggest using a larger *k* for better normals [Besl and Jain, 1988; Hoffman and Jain, 1987; Rabbani et al. 2006]. Yang and Feng [2005] point out that using a large number of points can adversely affect the local characteristics of the normal vector but the local geometry is better represented by a smaller number of points. Since the quality of the surface normals depend heavily on the structure of the surface

geometry it is better to investigate the problem empirically rather than analytically by simulation or related real data experimentation and find the best value for *k*. A neighborhood size should be sufficiently large so that neighbours in the local neighborhood become co-planar. Fig. (6) shows that for RPCA noise has lerss influence for neighborhood size and gives more consistent estimates (e.g. curvature) than PCA.
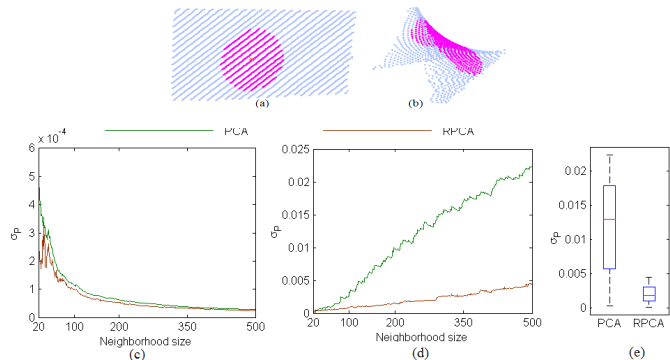

Figure 6.   (a) Planar surface, (b) non-planar smooth surface; magenta points are in a local neighborhood, k=500; line diagrams of $\sigma(p)$ vs k, (c) points from planar surface, (d) points from non-planar smooth surface, (e) box-plots for $\sigma(p)$ from non-planar smooth surface points

3) Curvature: Curvature is another criterion for proximity. It measures the rate of change of surface normal, which is fairly popular in point cloud data analysis. Many authors use different types of curvature measures from different methods in many ways for segmentation [Besl and Jain, 1988; Pauly et al. 2002; Rabbani et al. 2006]. Gaussian and mean curvature is proposed [Besl and Jain, 1988], but its limitations analysed [Powel et al. 1998; Rabbani et al. 2006] for its over segmentation and inefficiency even on very simple scenes with low noise. The residual with a percentile based cut-off value is proposed [Rabbani et al. 2006] as curvature, and Klasing [2009] discusses its limitations for segmentation. One of the limitations is that the residual is not normalized and its cut-off is fixed. We consider $\sigma(p)$ in Eq. (3) as the curvature and use it to determine the seed points for region growing. This normalized measure considers surface variations in all directions. It does not need extra computation and is available as a by-product of PCA. The problem is that PCA based $\sigma(p)$ is sensitive to outliers, and an alternative remedy is to use of RPCA [Nurunnabi et al. 2012]. Fig. 6(c, d) shows the RPCA based values of $\sigma(p)$ are more consistent than those that are PCA based, both for planar surfaces (Fig. 6(a)) and non- planar smooth surfaces (Fig. 6(b)). Box-plots in Fig. 6(e) show a comparative summary of robustness for PCA and RPCA based curvatures.

C. Algorithm Formulation

The algorithm proposed is based on region (surface) growing. It uses various local surface point proximity and/or coherence criteria. It begins with a seed points identified as the local surface points that have the least curvature $\sigma(p)$ values because it is expected that region growing will be more successful for areas where the surface variation is the least. We take a *k* neighborhood (NP_i) for the seed point p_i.

To make the surface smooth and at the same time to reduce outlier effects (if any), we use the point to plane Orthogonal Distance (OD) for all the neighbours of p_i that we can calculate at the time of normal estimation. The OD for the ith point p_i to its best fit plane generated by its neighbors can be defined as:

$$OD(p_i) = (p_i - \bar{p})^T . \hat{n} \qquad (6)$$

where $\bar{p}$ and $\hat{n}$ are the centre (mean) and the unit normal of the best fit plane. We follow the general rule to get the OD threshold (OD_{th}): (median (.) + a × MAD (.)) (a =1 to 3) for smooth co-surface points. To make the threshold robust we use the median and Median Absolute Deviation (MAD) instead of mean and standard deviation. We consider the ith point as a co-planar surface point if:

$$OD(p_i) < median\{OD(NP_i)\} + 2 \times MAD\{OD(NP_i)\} = OD_{th}, \quad (7)$$

where $\{OD(NP_i)\}$ is the set of all ODs for all the points in the neighborhood, and:

$$MAD = b.\underset{i}{median}| p_i - \underset{j}{median}(p_j)|, \qquad (8)$$

where b = 1.4826 to make the estimator consistent [Rousseeuw and Croux, 1993].

In region growing, two points defined to be spatially connected means that they should be locally close. Hence the distance between two points needs to be measured. They should be as close as the majority of the points that are close to each other in their neighborhood. We use the Euclidian Distance (ED) to compute the point to point distance (closeness) between a pair of points:

$$ED(p_{ij}) = || p_i - p_j ||, \qquad (9)$$

where p_i is the seed point and p_j is one of its neighbours. We consider p_j to be spatially connected to p_i if $ED(p_{ij}) < median\{ED(p_{ij})\} = ED_{th}$, where $\{ED(p_{ij})\}$ is the set of all EDs between the seed point and its neighbours, and ED_{th} is the ED threshold.

It is usually the case that an angle between two points from different feature surfaces (i.e. different local tangent planes) will be larger than a minimum angle threshold (θ_{th}), and on a smooth surface the angle between the two spatially close points will be much less. The angle (θ, sometimes termed the bias angle) between two points can be defined as:

$$\theta_{ij} = \arccos| \hat{n}_i^T \hat{n}_j |, \qquad (10)$$

where $\hat{n}_i$ and $\hat{n}_j$ are the two unit normals for the ith point and one of its neighbors p_j. To avoid the 180° ambiguity of the normal vectors, we use absolute values in Eq. (10). Two spatially close points will be co-surface points if θ is less than a predefined threshold θ_{th}.

Some of the proposed techniques in the literature use only θ, some use θ and curvature, and some suggest θ and ED for region growing in their segmentation algorithms. We argue all the three criteria (OD, ED and θ) are necessary for region growing. We illustrate this requirement by taking two datasets from two planar surfaces that have same orientation but parallel offsets (Fig. 7(a)). We add 20% noise to make the surface non smooth, Fig. 7(b). First we use only θ as a criterion to separate the planes but it fails, (Fig. 7(c)) because the surfaces have similar normals. The use of ED is necessary to make the surfaces different (Fig. 7(e)), and OD is necessary to make the surfaces smooth (Fig. 7(d)). Fig. 7(f) shows the advantage of using all three criteria for proper segmentation.
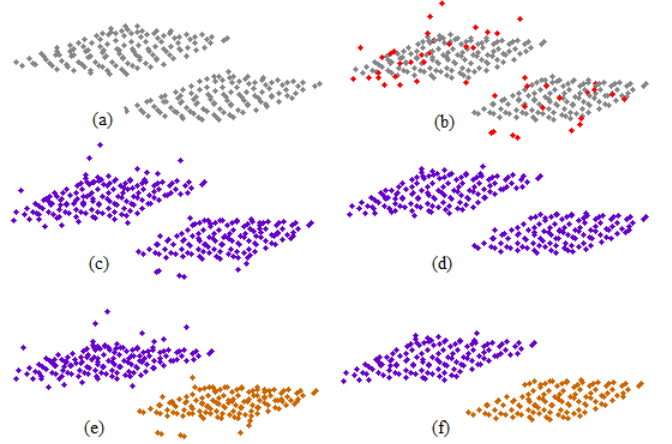

Figure 7. Region growing flow (a) step planar surface data (b) outlier (red points) contaminated data; robust segmentations based on (c) θ (d) θ and OD (e) θ and ED (f) θ, OD and ED, showing only those points that are determined to not be outliers

The region growing algorithm used is derived from the well known one based on seed points and iteratively growing each region by adding more points. After getting an initial seed point p_i and its neighbours for a current region R_c, we add a neighbour p_j to R_c and the current seed point list S_c with the conditions: (i) $OD(p_j) < OD_{th}$ (ii) $ED(p_{ij}) < ED_{th}$ and (iii) $\theta_{ij} < \theta_{th}$, and remove it from P (the point cloud). Now p_j will be considered as the next seed point for the R_c. The R_c will grow until any new seed point is available for R_c. If the size of R_c is less than a minimum number (R_{min}) of points then the region will be considered as insignificant. After getting a complete region, we select a seed point for the next region from the remaining points in P that has the least $\sigma(p)$. This process will continue until P is empty. The process is summarized below in Algorithm 1 and a diagram for the proposed robust segmentation is sketched in Fig. 8.
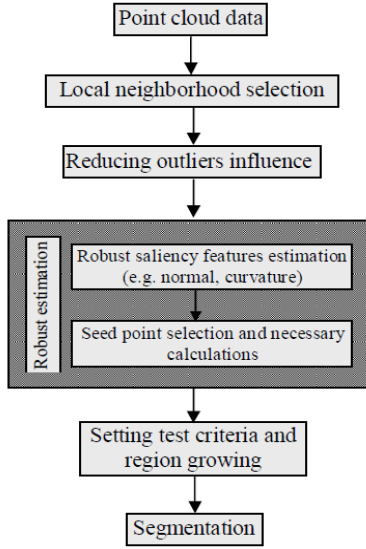
5

Figure 8. Robust segmentation diagram

Algorithm 1: Segmentation

Input:

Point cloud $P=\{p_i\}$, set of normals $N=\{\bar{n}_i\}$, set of curvature values $\{\sigma(p_i)\}$, set of $ODs=\{OD_i\}$, angle threshold θ_{th}, minimum region size R_{min}, current region $R_c=[\ ]$, current seed point list $S_c=[\ ]$, list of regions $R=[\ ]$.

Output:

   $R=\{R_c\} \leftarrow$ List of regions

while P is not empty **do**
   find initial seed point p_i from P with minimum $\sigma(p_i)$
   insert this seed point p_i into R_c and S_c
   for each seed point in S_c **do**
      find k-nearest neighborhood NP_i
      calculate ED_{ij} from seed point p_i to its neighbors p_j
      calculate ED_{th} using the EDs (Eq. (9)) of neighbouring points of NP_i
      calculate OD_{th} using ODs of neighbouring points NP_i using Eq. (7)
      calculate θ_{ij} between seed point p_i and its neighbour p_j using Eq. (10)
      for each neighbor in NP_i **do**
         if p_j is in P and $OD_j< OD_{th}$ and $ED_{ij}< ED_{th}$ and $\theta_{ij}<\theta_{th}$
         then
            insert p_j into R_c and S_c, and remove it from P
         end if
      end for
   end for
   if size of $R_c>R_{min}$ **then**
      insert R_c into R
   else
      the region is insignificant
   end if
   clear R_c and S_c
end while
sort the regions in R
return R

IV. EXPERIMENTS

The segmentation algorithm is demonstrated and evaluated in this section through experiments on three real laser scanning point cloud datasets. We compare the results in segmentation based on the covariance estimates using both of PCA and RPCA. We also see the neighborhood size and angle threshold have an effect on both PCA and RPCA segmentation.

A. Data set 1

This dataset contains 25,749 points of several type of roadside inventories (poles and signposts) (Fig. 9(a)). The dataset is a real laser scanning point cloud captured by a mobile mapping system. This data consists of a number of planar regions (road signs), long cylindrical and approximately cylindrical surfaces (poles) and more complex surfaces (lamps). We colour each data point to show the separate resultant segments determined by the algorithms. Based on preliminary experiments not reported here, we set the parameters $k=50$, $\theta^{th}=15°$ and $R_{min}=10$. Fig. 9(b), shows the result of PCA based segmentation. Over and under segmentation occurs in many regions. The results based on RPCA shown in Fig. 9(c), show almost perfect segmentation for the objects under study. The far right road sign shows some leakage of points from the two upright poles into the planar road sign. For the same road sign, the points describing the tops of the two poles protruding above the sign have been erroneously classed as outliers.
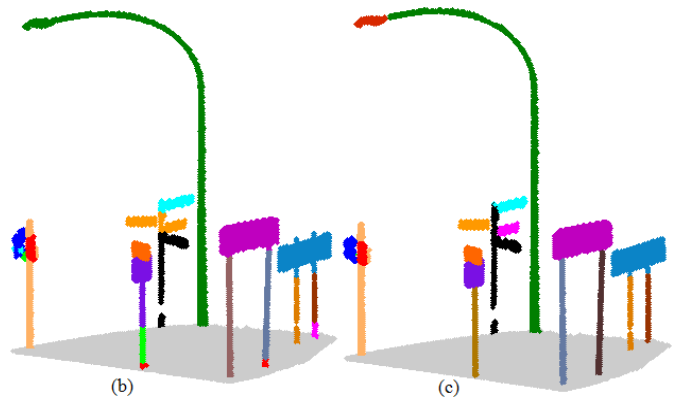

Figure 9. (a) Real point cloud data (b) PCA segmentation (c) RPCA segmentation,

6

B. Data Set 2

Fig. 10(a), this is a real point cloud data consisting of 12,769 points acquired from the mobile mapping system. It contains a streetlight pole (non-planar smooth surface) and a signpost (planar surface). To illustrate the performance in the presence of outliers, we add 2% (255) extra points with 0.2 normal variance are added. We set k=50, θ_{th} =18° and R_{min}=10. The PCA based result (Fig. 10(b)) shows that there are four cases of under segmentation and one case of over segmentation (pole is attached with the ground). Fig. 10(c) shows that RPCA segments the data properly into seven segments. To see how the neighborhood size change affects segmentation, we change the neighborhood size to k=40 and 60 and keep the same values for. θ_{th} and R_{min} as before. Fig. 10(d, e, f, g) show that for both neighborhood sizes 40 and 60, the results for RPCA are significantly better than for PCA. RPCA based segmentation is more consistent than PCA as a function of neighbourhood size although there are some less than perfect results for RPCA e.g. Fig. 10(g) in which one of the branches of the light has been separated from the rest of the lamp.
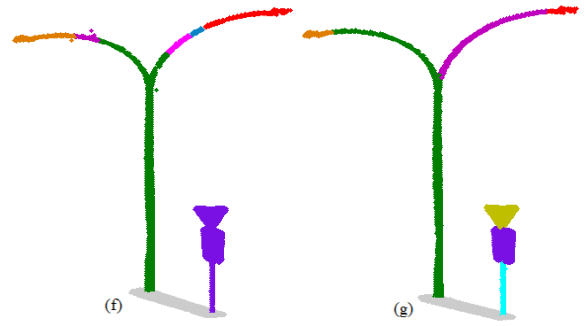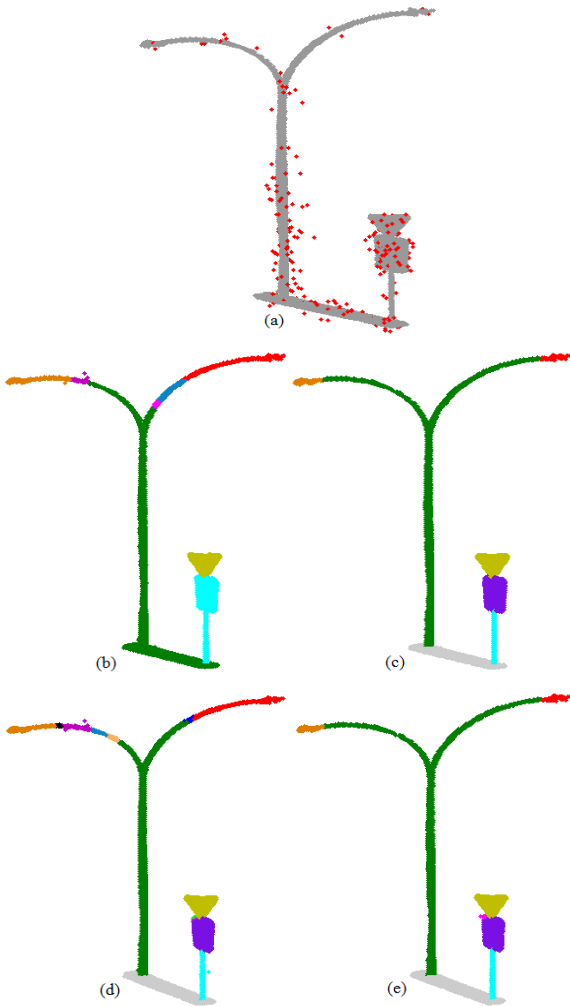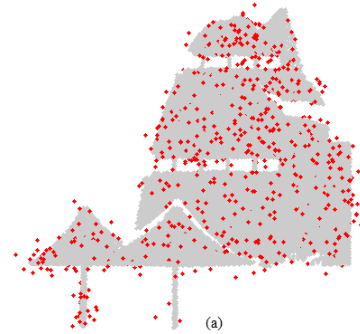

Figure 10. (a) Real point cloud data (b) PCA segmentation, k=50 (c) RPCA segmentation, k=50 (d) PCA segmentation, k=40 (e) RPCA segmentation, k=40 (f) PCA segmentation, k=60 (g) RPCA segmentation, k=60

C. Data set 3

This dataset (Fig. 11(a)) consists of 18,191 points for one road side building (shop), again captured from the mobile mapping system. We set k=30, θ_{th} =10°, and R_{min}=10. We add 5% outliers (noise). Fig. 11(b) shows the presence of over segmentation and under segmentation for PCA, whereas Fig. 11(c) shows perfect segmentation for RPCA. To see the effects of change in the angle threshold, we set θ_{th}=13° but still get unsatisfactory results for PCA segmentation (Fig. 11(d)). In Fig. 11(e), RPCA gives consistent segmentation and significantly better results than PCA. The different features around the windows as well as the vertical window bars have been separated from the main building and the umbrella poles have been segmented as complete features.
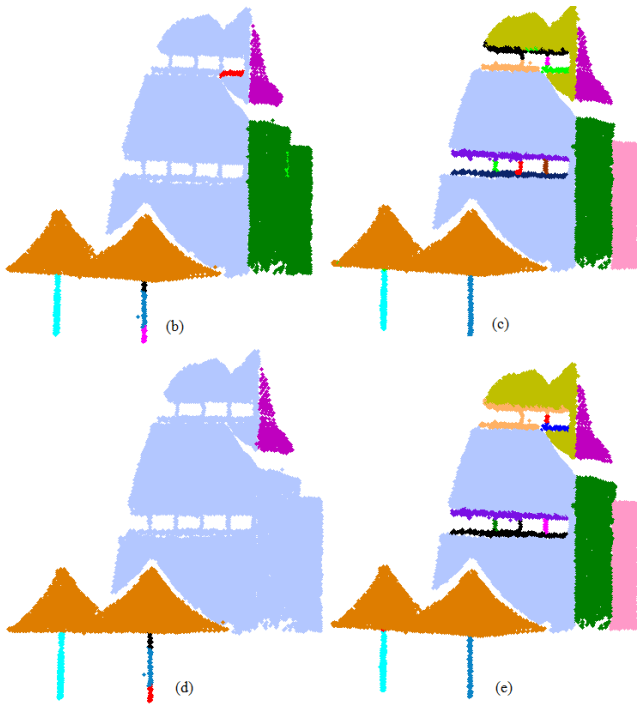
Figure 11. (a) Real point cloud data (b) PCA segmentation, θ^{th} =10°, (c) RPCA segmentation, θ^{th} =10°, (d) PCA segmentation, θ^{th} =13°, (e) RPCA segmentation, θ^{th} =13°

V. CONCLUSIONS

A region growing based statistically robust segmentation algorithm is proposed for planar and non-planar smooth surface extraction from laser scanning 3D point cloud data. The algorithm uses robust PCA, to make the saliency features used for region growing more resistance to outliers. Results for experiments on real mobile mapping point cloud datasets show that the proposed RPCA based algorithm outperforms PCA for segmentation producing more correct results. The RPCA results are more consistent over changes in neighborhood size and angle threshold. The algorithm breaks down at more than 50% outliers and the increase in outlier tolerance level and efficient segmentation for non-smooth surface are the subjects of future research.

REFERENCES

[1] P. J. Besl, and R. C. Jain, "Segmentation through variable- order surface fitting," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 10 (2), pp. 167–192, 1988.

[2] K. Klasing, "Surface-based segmentation of 3D range data," Technical report TR-LSE-2009-10-1, Institute of Automatic Control Engineering, Technische Universität München, Germany, 2009.

[3] J. Linag, F. Park, and H. Zhao, "Robust and efficient implicit surface reconstruction for point clouds based on convexified image segmentation," 2011.

[4] Y. Liu, and Y. Xiong, "Automatic segmentation of unorganized noisy point clouds based on the Gaussian map," Computer-Aided Design," vol. 40, pp. 576–594, 2008.

[5] M. Pauly, M. Gross, and L. P. Kobbelt, "Efficient simplification of point-sampled surface," in The Conference on Visualization, Washington, pp. 63–170, 2002.

[6] S. Sotoodeh, "Outlier detection in laser scanner point clouds," in IAPRS, Dresden, vol. XXXIV, pp. 297–301, 2006.

[7] T-J. Fan, G. Medioni, and R. Nevatia, "Segmented descriptions of 3-D surfaces," IEEE Journal of Robotics and Automtion, vol. RA-3, no. 6, pp. 527–538, 1987.

[8] H. Woo, E. Kang, S. Y. Wang, and K. H. Lee, "A new segmentation method for point cloud data," International Journal of Machine Tools and Manufacture, vol. 42, no. 2, pp. 167–178, 2002.

[9] X. Y. Jiang, H. Bunke and U. Meier, "Fast range image segmentation using high-level segmentation primitives," in WACV' 96: Proceeding of the 3rd IEEE Workshop on Applications of Computer Vision, IEEE Computer Society, Washington, DC, USA, p. 83, 1996.

[10] D. Marshall, G. Lukacs, and R. Martin, "Robust segmentation of primitives from range data in presence of geometric degeneracy," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 304–314, 2001.

[11] M. W. Powell, K. W. Bowyer, X. Jiang, and H. Bunke, "Comparising curve-surface range image segmenters," in Proceedings of the 6th International Conference on Computer Vision, IEEE computer society, Washington, DC, USA, p. 286, 1998.

[12] C. C. Chen, and I. Stamos, "Range image segmentation for modeling and object detection in urban scenes," in 3-D DIM, pp. 185–192, Quebec, Canada, 2007.

[13] H. Hoppe, T. De Rose, and T. Duchamp, "Surface reconstruction from unorganized points," in Proceedings of ACM SIGGRAPH, vol. 26, no. 2, pp. 71–78, 1992.

[14] N. J. Mitra, and A. Nguyen, "Estimating surface normals in noisy point cloud data," in 19th ACM Symposium on Computational Geometry, San Diego, California, pp. 322–328, 2003.

[15] B. Li, R. Schnabel, R. Klein, Z. Cheng, G. Dang, and S. Jin, "Robust normal estimation for point clouds with sharp features," Computers and Graphics, vol. 34, pp. 94–106, 2010.

[16] M. Pauly, R. Keiser, and M. Gross, "Multi-scale feature extraction on point-sampled surfaces," Eurographics, vol. 22, no. 3, 2003.

[17] T. Rabbani, F. A. van den Heuvel, and G. Vosselman, "Segmentation of point clouds using smoothness constraint," in IAPRS, XXXVI (5): pp. 248–253, 2006.

[18] B. Oehler, J. Stueckler, J. Welle, D. Schulz, and S. Behnke, "Efficient multi-resolution plane segmentation of 3D point clouds," in 4th ICIRA, Achen, Germany, 2011.

[19] R. A. Maronna, and V. Yohai, "Robust estimation of multivariate location and scatter," Encyclopedia of Statistics, vol. 2. New York: John Wiley, 1998.

[20] M. Hubert, and P. J. Rousseeuw, "ROBPCA: A new approach to robust principal component analysis," Technometrics, vol. 47, no. 1, pp. 64–79, 2005.

[21] P. J. Rousseeuw, and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," Technometrics, vol. 41, no. 3, pp. 212–223, 1999.

[22] R. A. Maronna, and V. J. Yohai, "The behavior of the Stahel-Donoho robust multivariate estimator," Journal of the American statistical association, vol. 90, no. 429, pp. 330–341, 1995.

[23] J. Friedman, and J. Tukey, "A projection-pursuit algorithm for exploratory data analysis," IEEE Transactions on Computers, vol. 23, pp. 881–889, 1974.

[24] W. A. Stahel, "Robust estimation: infinitesimal optimality and covariance matrix estimators," PhD Thesis, ETH, Zurich, 1981.

[25] D. L. Donoho, "Breakdown properties of multivariate location estimators," PhD qualifying paper, Harvard University, 1982.

[26] M. Yoon, Y. Lee, S. Lee, I. Ivrissimtzis, and H-P. Seidel, "Surface and normal ensembles for surface reconstruction," Computer-Aided Design, vol. 39, pp. 408–420, 2007.

[27] R. Hoffman, and A. K. Jain, "Segmentation and classification of range images," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-9, no. 5, pp. 608–620, 1987.

[28] C. Wang, H. Tanahashi, and H. Hirayu, "Comparison of local plane fitting methods for range data," in IEEE Computer Society Conference, CVPR'1, Kauai, vol. 1, pp. 663–669, 2001.

[29] D. O. Yang, and H-Y. Feng, "On normal vector estimation for point cloud data from smooth surfaces," Computer-Aided Design, vol. 37, pp. 1071–1079, 2005.

[30] K. Klasing, D. Wollherr, and M. Buss, "Realtime segmentation of range data using continuous nearest neighbors," In Proceedings of the IEEE International Conference on Robotics Automation, 2009.

[31] K. D. Tamal, L. Gang, and J. Sun, "Normal estimation for point cloud: a comparison study for a voronoi based method," Eurographics Symposium on Point-Based Graphics. M. Pauly, M. Zwicker, (Editors), pp. 39–46, 2005.

[32] A. Nurunnabi, D. Belton, and G. West, "Diagnostic-robust statistical analysis for local surface fitting in 3d point cloud data," in Proceedings of ISPRS Congress, Melbourne, Australia, in press, 2012.

[33] P. J. Rousseeuw, and C. Croux, "Alternative to the median absolute deviation," Journal of the American Statistical Association, vol. 88(424), pp. 1273–1283, 1993.