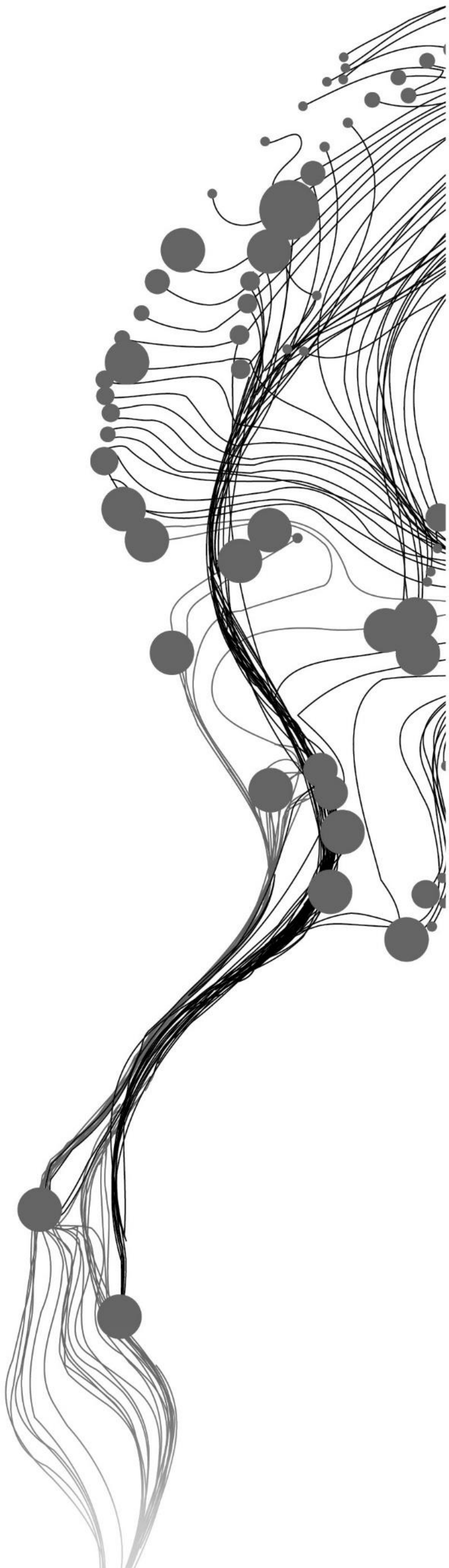
THE INFLUENCE OF POINT REDUCTIONON THE SEGMENTATION OF MLS DATA

Wanpeng Shao

March, 2016

SUPERVISORS:
Dr. M. S. Peter
Prof. Dr. Ir. M.G. Vosselman



THE INFLUENCE OF POINT REDUCTION ON THE SEGMENTATION OF MLS DATA

Wanpeng Shao

Enschede, The Netherlands, March, 2016

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Dr. M. S. Peter
Prof. Dr. Ir. M.G. Vosselman

THESIS ASSESSMENT BOARD:
Prof. Dr. Ir. Alfred Stein (Chair)
Dr. R. C. Lindenbergh (External Examiner, Delft University of Technology, Optical and Laser Remote Sensing)

ABSTRACT

Laser scanning is developing rapidly, which can measure millions of points with very high accuracy coordinates of measured surfaces, a large set of data can provide a wealth of information of surveyed surfaces. MLS is one of the important categories of the laser scanning systems, which have been put into use in various fields like highway maintenance, survey structural analysis and city modelling etc. Various mobile laser scanning systems are produced so as to fulfil users' needs. While different MLS specifications and movement of platforms lead to uneven point distributions, which make it difficult to come up with a proper neighborhood definition to extract features, especially for the surface growing segmentation. So one of this research objectives is to analyze the effects of point distributions on segmentation results of MLS data.

Actually we do not need to put all MLS data into applications. Since a large number of points will cause huge computational cost. While the way of point reduction can reduce points and change point distributions, which have a big impact on segmentation results. So the other objective of this research is to analyze the effects of point reduction on the segmentation of MLS data.

Here two quality assessment methods are combined together to assess the quality of segmentation results. One of methods is visual assessment, which means segmentation results are inspected visually. The other one is quantitative analysis based on the point level. Values of measures can be computed by quality assessment model. Evaluation of every point reduction's applicability is done according to segmentation accuracy, users can select the most suitable one according to the specific application.

To sum up, surface growing segmentation is used to extract features from point clouds in this study. Two datasets acquired by different scanning configurations are used as research data. One data set was recorded in Enschede, and the other one was recorded in Paris. Six point reduction methods existed in mapping library and CloudCompare "subsample" tool aim at removing points from two research datasets. In terms of effects of point distribution on segmentation, analysis is only based on segments in one façade dataset, segmentation accuracy of façade data in Paris data and Enschede data is 0.51 and 0.914 respectively. The reason is that Enschede data is more even than Paris data. In terms of effects of point reduction on segmentation, only one façade dataset is used as analyzed data. For Paris data, there exists two cases: one of cases is that after point reduction, segmentation accuracy of the façade data reach to 0.845, the reason is that its point distribution is become more even than before. While the other case is that the segmentation accuracy fall to 0.368, the reason is that point distribution is still uneven and even worse than original data. But for Enschede data, segmentation accuracy of all reduced façade datasets are very stable with high accuracy , because resulting point distribution is still even, so all methods has good applicability used on Enschede data. In terms of evaluation of applicability of these six point reduction methods on Paris data, considering segmentation accuracy of façade data and their stability, point reduction based on "space" is selected as the most efficient one.

Keywords: mobile laser scanner, point cloud, point reduction methods, visual assessment, quality assessment model, applicability

Acknowledgements

I would like to take this opportunity to express my sincere thankfulness and appreciations to my first supervisor Michael Peter and my second supervisor prof.dr.ir. M.G. Vosselman (George) who gave me the golden advice to guide me to finish my thesis. They helped me in doing a lot of Research and I came to know about so many new things. I am really thankful to them. I also would like to thank Yizhen, Lao, Yancheng Wang for providing me some suggestions about programming guidance, dr. B Xiong (Biao) and PHD Fashuai Li for giving me important guidance of my research.

Secondly, I would also like to express my gratitude to my friends, who give me life and academic help during the 18 months study life in the Netherlands: Na Ji. Jiawei Zhang, Jianwei Pan, Jiatan Jiang. They give me great courage to accomplish my research. Absolutely, I really thank my family. Their love and caring support me to finish the study in the Netherlands.

In the end, thanks to all staff of Earth of Observation Science department of ITC. The study experience is very important in my life. Without their constant help, I will not make it.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Motivation and problem statement

Laser scanning technique plays an important role in surveying and mapping in recent years. Accordingly, laser scanning technique is developing rapidly which has been put into use in various fields such as Mining (Khoshelham, et al. 2011), Civil Engineering and Construction(Gonzalez-Jorge, et al. 2012), etc. In addition, point cloud obtained by laser scanners is beneficial for knowledge based recognition of objects (Pu, et al. 2011).

Laser scanning devices enable to measure numerous number of points from objects with high accuracy (Moenning & Dodgson, 2003). For instance, commercial laser scanners installed on the car can reach objects which is more than 500 m away from the platform, and produce 550000 pts/s with only 5 cm accuracy (Puente, et al. 2013). Such high point density can provide a wealth information of measured objects.

However, the computation cost poses a big challenge for processing a large set of point clouds. And a large amount of redundant points included in original point cloud which should be removed. For example, in the case of visualization of 3D objects, like car, sign, façade, ground, not so many points are needed to describe them. So it is a meaningful thing to remove these redundancies.

Mobile laser scanning is one of the important categories of laser scanning technique, a variety of MLS systems have been described in(Puente, et al. 2013). Different scanning configurations and instability of speed of platform will result in different point distributions, point clouds with different point distributions are uneven (Rutzinger, et al. 2010). Uneven point distribution makes it difficult to come up with an appropriate neighborhood definition used for feature extraction (Weinmann, et al. 2015), especially for surface growing segmentation. While, there is not too much research on analyzing the impacts of different point distributions on segmentation of MLS data.

While point reduction can change point distribution, even make it more even than initial dataset. The reduced dataset with even point distribution is helpful to segment point cloud. There exists some point reduction methods in Mapping library and CloudCompare, which can be applied to decrease point density of point clouds. However, no researchers focus on studying the influence of these existing point reduction methods on segmentation of MLS data. So this study aim at finding a way to investigate the effects of point distributions and point reduction on surface growing segmentation of MLS data.

1.2 Research identification

1.2.1 Research objective

The overall objective of this study is to evaluate the influence of different point distributions and existing point reduction methods on surface growing segmentation. The way of evaluating these point reduction's applicability is also defined in this study. So objective of this research is described as follows:

I. Find a way of investigating the influences of point distributions /point reduction on the surface growing segmentation of MLS data.
II. Find out how to evaluate these existing point reduction methods' applicability on MLS data.

1.2.2 Research questions

In order to achieve two objectives of research. The following questions should be solved:

I. What is the difference between point clouds acquired by mobile laser scanning systems with different configurations?
II. What is the effect of point distribution on the segmentation of MLS data?
III. What is the influence of point reduction to the segmentation results in terms of quality?
IV. How can the quality of the resulting segmentation be assessed (with respect to the point distribution/reduction)
V. Evaluate applicability of every point reduction method and to see whether to propose a point reduction method to optimize segmentation algorithm?

1.2.3 Innovation

In order to put point clouds into a variety of applications, various mobile laser scanning systems are produced(Puente et al. 2013). These MLS systems have different scanning configurations, which result in different point distributions. But almost no researchers aim at analyzing influence of point clouds obtained by these MLS systems on surface growing segmentation. In addition, point reduction is an efficient way to reduce redundant points and change its point distribution (Lee, 2009), which has a big impact on segmentation accuracy. There has existed some useful point reduction methods in Mapping library and CloudCompare. However, there is no research on investigating what impact of these point reduction methods have on segmentation of MLS data. And analysis should be based on segmentation accuracy, but there is not much research about assessing quality of segmentation results, so a new quality assessment model proposed in this research is applied to assess segmentation accuracy. Overall, innovations of this research are as follows:

I. Effects of different existing point reduction/distribution on segmentation methods of MLS data are analyzed in this study.
II. A novel quality assessment model to evaluate accuracy of segmentation results is proposed in this research

1.3 Thesis Structure

Chapter 1 introduces the background of this study containing motivation, problem statement and research identification. Various MLS systems, segmentation methods, a variety of point reduction strategies, and a quality assessment method are reviewed in chapter 2. Chapter 3 describes parameter settings in surface growing segmentation, six existing point reduction methods, quality assessment model, and the way of analyzing point reduction\point distribution on segmentation results. Chapter 4 shows analysis on segmentation results from original and reduced datasets. In Chapter 5, some problems and limitations of this study are discussed. Chapter 6 presents answers to every research questions and further developments.

2 Literature Review

2.1 Introduction

Literature review is divided into 4 parts. First part describes various MLS systems. A variety of segmentation methods are briefly reviewed in part 2. And part 3 shows three point reduction methods. There are not many studies on assessing segmentation accuracy of point clouds, so only one quality assessment method is presented in final part.

Various different MLS systems are reviewed in (Puente, et al. 2013), such as ROAD SCANNER-SITECO, IP-S2-TOPCON, MX8-TRIMLE, VMX-250-RIEGL, etc.

Surface growing segmentation, connected component segmentation, and scan line segmentation are described in (Vosselman, et al. 2004). And (Vosselman, 2013) describes segment growing segmentation, and majority filtering, and mean-shift segmentation in detail.

Only one quality assessment method is reviewed to assess segmentation accuracy. The method described in (Ni, et al. 2014) is to transform 3D point cloud into 2D image, but space and color information is maintained, then use the 2D image to do segmentation. Ground truth segments is selected by visual interpretation. After that, compare segmentation results with ground truths according to several ways (area-based, location-based, and combined measured), and calculate values of measures (Over-Segmentation, Under-Segmentation, RPsuper, D, M). Over and under segmentation is area-based. RPsuper means relative position between centroid of ground truth segment and compared segment. D and M are combined measures, D combines under-segmentation and over-segmentation and M consists of area-based and location-based measures. Even though this method is more reliable than subjective evaluation, the ground truths have to be selected manually. In this study, measured objects are quite complex, like car. If user transform MLS point cloud to 2D image, 3D information contents of objects will be lost. Evaluation of façade point cloud segmentation of objects like car (along the projection direction) is not accurate. So it is not well-used for assessing segmentation accuracy of MLS data.

2.2 Review of Mobile laser scanner systems

(Puente, et al. 2013) comprehensively describes specifications and configurations of the latest MLS systems. This review focuses on two factors: MLS configurations and specifications. MLS specifications include range precision, range accuracy, scan frequency, field of view, maximum range, and measurement per laser pulse. MLS configuration means the way of laser scanners installed on the platform. A variety of MLS systems is shown in Figure 2-1 and specifications of every MLS system are shown in Table 2-1.

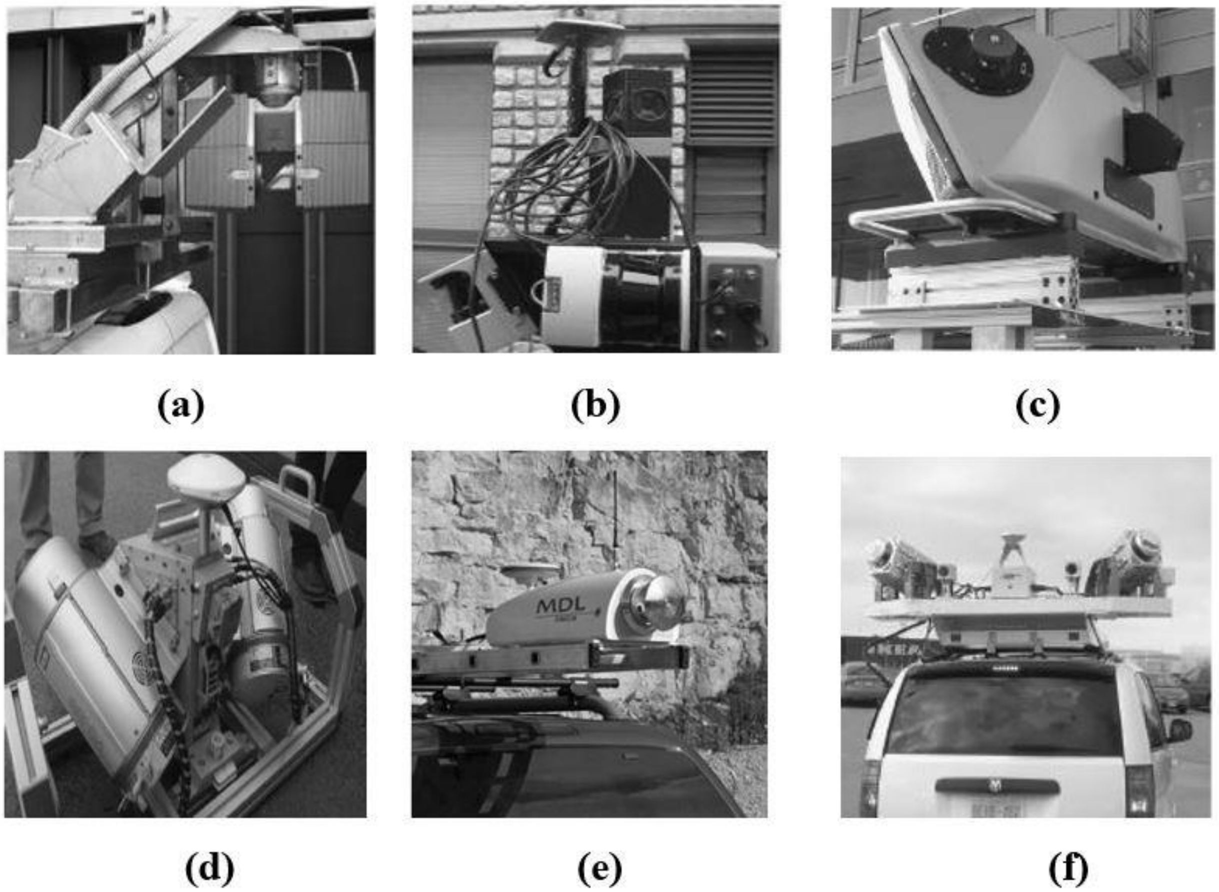


Figure 2-1(a) FARO PHOTON 120 (SITECO ROAD SCANNER) (b) TOPCON IP-S2 SICK laser scanners (c) TRIMBLE MX8 scanner (d) STREETMAPPER RIEGL VMX-250 (e) MDL DYNASCAN Single MDL Laser scanner (f) OPTECH LYNX MOBILE MAPPER (Puente et al. 2013)

Scanner	FARO Photon 120	SICK LMS 291	VQ-250	MDL scanner	LYNX laser scanner
MLS system	ROAD SCANNER	IP-S2	VMX-250	DYNASCAN	LYNX MOBILE MAPPER
Maximum range	120m	30m	200m	More than 500m	200m
Range accuracy	± 2 mm	± 35 mm	$+10$ mm	± 5 mm	± 10 mm
Laser measurement rate	122-976 kHz	40 kHz	550 kHz	36 kHz	75-500kHz
Scan Frequency	48Hz	75Hz	Up to 100Hz	Up to 30Hz	80-200Hz
Field of view	$\mathrm{H}360^0/V360^0$	$\mathrm{H}180^0$ $/V90^0$	360^0	360^0	360^0

Table 2-1 Specifications of MLS systems (Puente et al. 2013)

As we can see from Figure 2-1, A FARO PHOTON 120 is installed on the roof of the vehicle, which scans orthogonally to the driving direction; such that the side information of objects can't be obtained. Generally, three SICK LMS scanners are applied to obtain point cloud together, one of these point forward (or backward) to the driving direction, and others point to left and right side of the vehicle, but it is still impossible to get side information of measured objects. So the next generation products are installed non-orthogonally to the driving direction so as to get a complete point cloud of objects such as TRIMBLE MX8, RIEGL VMX-250, and LYNX MOBILE MAPPER.

From the Table 2-1, the most of accurate point cloud we can get is from FARO Photon, because it is based on phase difference technology. And the maximum range is MDL scanner, which can reach 500m objects away from the platform with around ± 5 mm accuracy. Every MLS has its own specification, so users can select the most suitable one according to the related application.

2.3 Review of segmentation methods

(Vosselman, et al. 2004) describes three different segmentation methods: Scan line segmentation, surface growing segmentation, and connected components in voxel space.

Scan line segmentation is divided into 2 phases: split and merge. In split phase, scan lines are split into line segments until the perpendicular distance from point to its corresponding line segment is below the threshold. The merge phase is performed in the region growing. In order to do region growing, seed region should be found at first. Seed region defined as 3 neighboring line segments that should satisfy three conditions: minimum overlap, a minimum length, and a maximum distance from neighbor points to two neighboring scan lines. Region growing is to merge other line segments if the perpendicular distance between two end points of additional adjacent line segments and estimated surface is below threshold. The key value is minimum region size, which highly determine success of a surface detection. One advantage of this method is that it can be used for detecting planar surface correctly. Disadvantage of this method is that it can't be used for extracting complicated shapes, such as cylinder objects. In this paper, scan line segmentation is applied to extract DEM in urban area. Because it can eliminate the bridge and fly-overs and tunnels correctly, which are connected to ground.

Surface growing segmentation is divided into two steps, seed selection and growing phase. In seed selection phase, Hough transform is used for finding seed surfaces, the key is size of points around seed point should be defined carefully. And in growing phase, there are three different criteria we can define: Proximity of points, locally planar, and smooth normal vector. For proximity of points, this criteria is usually based on k-nearest neighbors as extracted from a KD tree. For locally planar, this proximity can be applied by checking if the distance between a candidate point and the estimated plane is below threshold. Threshold of maximum distance to the plane and number of neighbors around a seed point determine smoothness of the resulting surface. For smooth normal vector field, this value is to check if the angle between normal of estimated surface and the normal of local surface generated from points around the seed is below the threshold. In this paper this method is used for detecting cylinders and planes segment in industry.

Connected components in voxel space is also divided into two stages. Firstly, all 3D points obtained from the measured object are converted into 3-dimensional voxels. Secondly, every voxel is labeled with the number, for example: voxels without any points are set to 0, voxels with points are labeled 1. Or the number of points within each voxel is counted, then assign the number to each corresponding voxel. The benefit of this method is some operations such as filter kernels and mathematical morphology can be used. In this paper, this method is applied to segment point cloud within trees, which can estimate the trunk and branch correctly.

segment growing segmentation, mean shift segmentation and majority filtering are described in (Vosselman, 2013). Segment growing segmentation is to determine a better spatial separability of neighboring segments. This method is well-used for point clouds without big planar segments, which is divided into two stages: seed detection and growing phase.

In seed detection phase, seeds with similar feature values like echo widths or normal vectors are detected. In growing phase, if the candidate point's feature value is close to average feature values of a segment, the seed are extended. But it can't detect real segments like surface growing does.

After segment growing segmentation, still a certain number of points are outliers without any segment number. So majority filtering can be helpful to assign these outliers into segments with the most frequent segment number around these isolated points.

Mean shift segmentation is mainly based on feature space like color, coordinates, and amplitude and pulse. This algorithm is mainly divided into two steps: Density estimation and Mode finding. It is mainly to find every point's centroid of cluster in feature space. Then points with the same features as centroid of cluster have are classified as one segment. For MLS dataset, point distribution is relative even, segments generated by mean shift segmentation method on the basis of point distribution will be inhomogeneous.

2.4 Review of quality assessment methods

A strategy from (Ni et al. 2014) aim at evaluating the segmentation accuracy in imagery and 3D laser point cloud data to find an optimal segmentation method and related parameters. For this purpose, a quality assessment model is described in this paper, which is based on criteria including differences in area and location, between ground truth segments and segments after segmentation. In this model, ground truth segments are manually selected. Segmentation results are regarded as compared ones. In the matching phase, here three structure is created: every point contains related segment ID and screen coordinate (x, y), every segment stores point with the same ID, matching is based on coordinates. Then put all points into the segment structure. After that, every ground truth segment are assigned to match resulting segments, if the number of intersected points in ground truth segment and reference segment is more than 20 pixels, these two related segments are matched. Finally, calculate values of core measures.

Area-based measures(Möller, et al. 2007) are applied to show differences of size between ground truth segments and segmentation results.

$$RAsub_{ij} = \frac{area(X_i \cap Y_j)}{area(X_i)}$$

$$RAsuper_{ij} = \frac{area(X_i \cap Y_j)}{area(Y_j)}$$

$\mathrm{X} = \{X_i\colon \mathrm{i} = 1 \dots \mathrm{n}\}$ is a set of ground truth segments and $\mathrm{Y} = \{y_j\colon \mathrm{j} = 1 \dots \mathrm{m}\}$ is a set of compared segments. $Area(X_i \cap Y_j)$ is the overlapping area of ground truth segments X_i and resulting segments Y_j. For these two values, 1 means the optimal match. Then values of over-segmentation and under-segmentation(Clinton, et al. 2010) can be calculated based on $RAsub_{ij}$ and $RAsuper_{ij}$.

$$\mathrm{OverSegmentation} = 1 - \frac{area(X_i \cap Y_j)}{area(X_i)}$$

$$\mathrm{UnderSegmentation} = 1 - \frac{area(X_i \cap Y_j)}{area(Y_j)}$$

These two values are continuous in [0, 1] with 0 being an optimal match.

Location-based measures define by (Möller, et al. 2007) describe relative distance between centroid of a ground truth segment and the related resulting segment.

$$RPsub_{ij} = \text{dist}\Big(\text{centroid}(X_i), \text{centroid}(Y_j)\Big).$$

$$RPsuper_{ij} = \frac{dis(centroid(X_i), centroid(Y_j))}{dist_{max} = max_i(RPsub_{ij})}$$

$RPsub_{ij}$ and $RPsuper_{ij}$are in [0, 1]. The lower values we calculated, the optimal segmentation result we can get. Then combined measures are calculated based on Over-Segmentation and Under-Segmentation, the root square(Weidner, 2008):

$$D_{ij} = \sqrt{\frac{OverSegmentation_{ij}^2 + UnderSegmentation_{ij}^2}{2}}$$

D is in [0, 1] with lower values means better match.

Another combined values M (Möller, et al. 2007) are also calculated, which is composed of area-based and location-based values.

$$M_{ij} = \sqrt{\frac{(1 - RAsub_{ij})^2 + (1 - RAsuper_{ij})^2 + {RPsub_{ij}}^2 + {RPsuper_{ij}}^2}{4}}$$

This value is in [0, 1] with 0 indicates perfect match.

Though 8 core measures are calculated in this paper to evaluate the segmentation methods, point cloud data used in this paper has to be transformed into 2D image at first. However, this method can only be used for data containing large studying area.

3 Method adopted

This study aim at investigating effects of point reduction on segmentation of MLS data. Since various mobile laser scanner systems with different configurations are already put into use, which result to different point distributions. In this study, two sets of point clouds are used so as to analyze point distributions on segmentation results. And there exists some point reduction methods in thinlaser tool, CloudCompare "subsample" tool, which can be applied to reduce MLS data. So this study is mainly to analyze effects influence of these point reduction methods on segmentation of two sets of MLS data. Two quality assessment methods are combined together to do analysis.

One of these is visual check, which means resulting segments are inspected visually. The other one is quantitative analysis according to quality assessment model, values of core measures can be calculated by quality assessment model. After analysis, applicability of every existing reduction method used in this study is evaluated.

In this chapter, surface growing segmentation is described in section3.1. Existing point reduction methods used in this study are explained in section 3.2. Section 3.3 introduces how to build quality assessment model. Section 3.4 and 3.5 present how to analyze influences of point distribution/point reduction on segmentation of MLS data and how to evaluate applicability of every point reduction method used in this study. The overall workflow applied in this study is shown as follow:

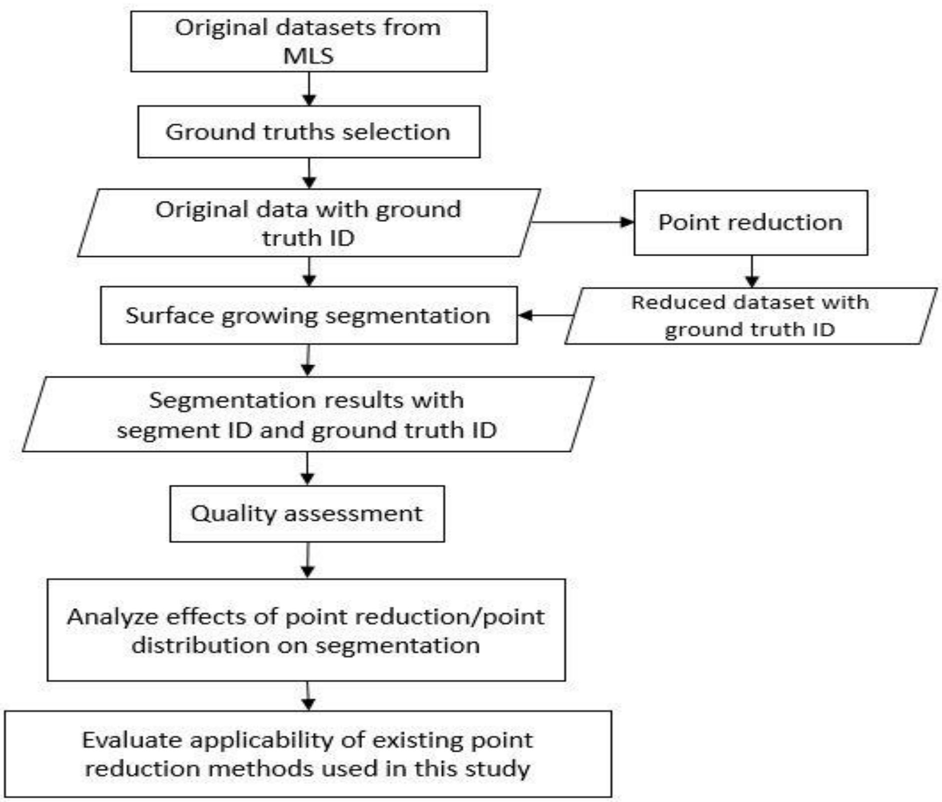


Figure 3-1 Workflow used in this research.

3.1 Surface growing segmentation

Surface growing segmentation method is applied to process all original and reduced datasets in this research, which is described in section 2.3. This chapter briefly describes some important parameters of surface growing segmentation.

In "Neighborhood definitions", there are 3 options in the storage model: TIN, Octree, and kd-tree. In this research, we only consider the kd-tree in this research. "Number of neighbors in kd-tree" specifies the number of neighboring points around the seed point, which is regarded as the most important variable in this research. This value is set to 20 at first because of high point density of MLS data.

"Surface growing phase" are divided into two steps: determination of seed surfaces and growing phase.

The seed surface determination is based on Hough transform. Two parameters highly affect seed surface determination, "Seed neighborhood radius" and "Maximum distance to plane". In case of high point density of MLS data, both "neighborhood radius" and "the number of neighbors in kd-tree" are used to restrict the neighboring points, the default value of "Seed neighborhood radius" is set to 1m. "Maximum distance to plane" specifies whether candidate points are assigned to the local detected plane. In this study, this value setting should be a trade-off of segmentation results between different features in the point cloud. Other parameters are left default values, only change the value of "Maximum distance to plane". When "Maximum distance to plane" is set to 0.2, segments in windows, wall, roof in Figure 3-2 is distinguishable by visual interpretation.

(a) Paris data (b) Enschede data

Figure 3-2 visualization of resulting segments of various objects

Once a seed surface has been determined, the growing phase is to find neighboring points and to check whether it is accepted into the same plane.

In this study, the surface model is "planar", PCM also have offered options to extract smooth. For the time being, only left it at "planar".

In growing phase, "Maximum distance to surface" also specify whether the candidate are assigned to a plane. If the distance from the candidate point to the fitted plane is below this value, this candidate point will be accepted, vice versa. In case of avoiding under-segmentation and over-segmentation, this value is also set to 0.2m.

3.2 Point reduction methods

Six existing point reduction methods are applied to reduce point clouds in this research: Method 1, method 2, method 3, method 4, and method 5 are implemented in ITC mapping library "thinlaser" tool. Method 6 exists in CloudCompare "subsample" tool.

Method 1 is based on "space", the distance between every two points should be more than a minimal distance defined by users. Here are one parameter in this method: min_distance. Once value of "min_distance" is set by user, 3D neighbourhood around the selected point will be generated, then other points within the range of selected point are deleted. So the larger value of distance set by users, the more points are deleted. Here if the distance is set to 0, no points are deleted.

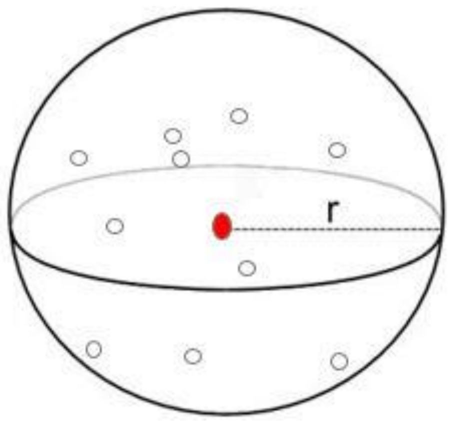


Figure 3-3 selected point (red) and unselected points (white)

Method 2 is to reduce point cloud by taking every n'th point. Three parameters exist in this method: reduction_factor, random and start_offset. In this method, value of "random" is set to 0, such that selected points are nonrandom. Offset of the first point is also left default value: 0, so this method starts from the first point in the whole dataset. Reduction factor is set by user, once reduction factor is determined, the whole dataset will be divided into the number of sets. From figure 3-4, if reduction factor is set to 3, only every first point of each group of "reduction_factor" point set is selected, so the results shows this method is to remove points by taking every 3'th point. But when reduction factor is set to 1, no points are deleted.

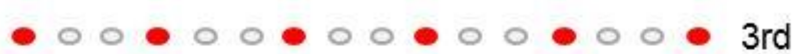


Figure 3-4 selected points (red) and unselected points (white)

Method 3 is to reduce data by randomly selecting one point from every set n points. Three parameters exist in this method: reduction_factor, random and start_offset. In this method, value of "random" is set 1, offset of the first selected point and is also selected randomly out of the first group of "reduction_factor" points. Then randomly select every point out of each group of "reduction_factor" points from the last selected point. From figure 3-5, the "reduction_factor" is set to 4, then the value of offset of the first points is randomly generated: 3, so the 3'th point is selected the first selected point. After that, randomly select one point out of the first group of 4 points from the 3'th point, here randomly value is 3, so the third point from the last selected is selected, loop this procedure within the whole dataset. So every time using this method with the same reduction parameter to reduce the same dataset, the resulting point distributions are not the same. For the time being, only one group of resulting datasets is selected to do further research.

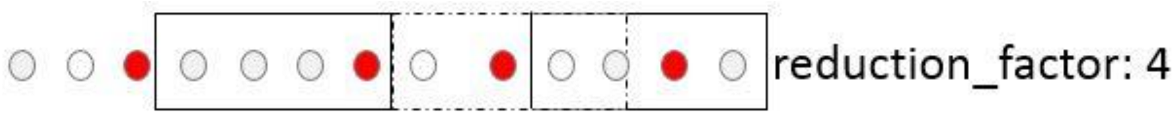


Figure 3-5 selected points (red) and unselected points (white)

Method 4 is to reduce points by taking every n'th point based on knn. There are three variables can be set: knn_start, knn_max, optimal reduction factor. knn_start and knn_max are left the default values. Here knn_start is 2, knn_max is 10. Optimal reduction factor can determine its reduction levels. Here the value of n is selected sequentially from knn_start to knn_max. Once one value of "n" is determined, the whole dataset will be divided into the number of sets, the seed point is selected. Specially, the selection of seed starting points in this method is random. Even though every time running this method with the same reduction parameter to reduce the same dataset, the resulting point distributions are different. Then reduction factor is calculated as follows:

$$\text{reduction factor} = \frac{the\ number\ of\ points\ within\ whole\ data}{the\ number\ of\ selected\ points}$$

If the reduction factor is smaller than optimal reduction value set by users, the value of "n+1" based on knn is used to do the same procedure. Loop until reduction factor is greater than or equal to optimal reduction factor. Finally, delete other unselected points. So only one group of resulting reduced data is used to do further research. If optimal reduction factor is set to 1, no points are deleted. For instance, if optimal reduction factor is set to 10, only the value of knn between knn_start and knn_max is selected to 10, calculated reduction factor will be equal to the optimal reduction factor. The loop will end. Such that only one point of each group of 10 points will be kept, which is shown in Figure 3-6.

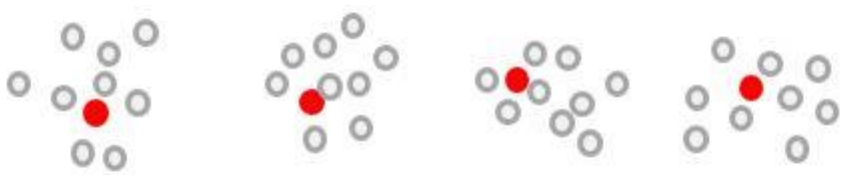


Figure 3-6 selected points (red) and unselected points (white)

Method 5 is to filter data by specifying minimum height difference between two points on the same XY location. This method is well used for point cloud obtained by Lynx scanning system. Lynx scanning system has two same Lynx scanners to acquire point clouds of measured objects, which will result in some objects scanned by two scanners. This method is mainly used to delete double points, before deleting the double point, Z coordinate of each double points is calculated, then assigned it to the selected point.

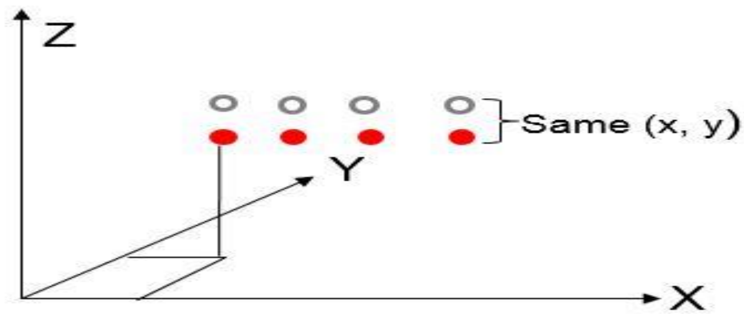


Figure 3-7 selected points (red) and unselected points (white)

Method 6 is to reduce points based on "octree", a level of subdivision of the octree is set by users. Once the subdivision level is set, the size of every small "Octree" is determined. For all points within "octree", only the nearest point to the octree center is kept. In this method, the number of deleted points is inverse proportional to the value of subdivision level.

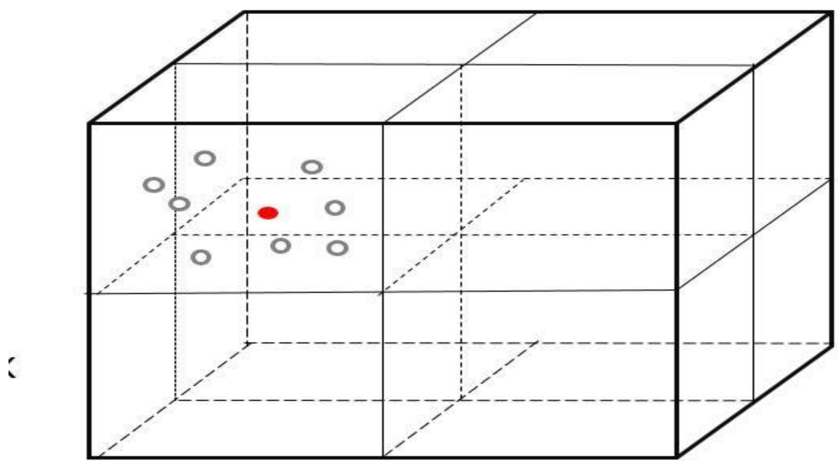


Figure 3-8 selected points (red) and unselected points (white)

3.3 Quality assessment Method

Two quality assessment methods can assess segmentation accuracy. First one is quality assessment model, values of core measures can be calculated by this model. The second is visual check, under-segmentation and over-segmentation can be visually inspected by users.

3.3.1 Quality assessment model

Figure 3-9 represents the workflow of quality assessment model. This algorithm consist two steps: first step here is called preprocessing stage, and the second one is calculation stage.

In the first stage, ground truth segments are selected manually. Ground truth selection phase is based on structure of object so as to make ground truths reliable. While selecting ground truths, points within the related ground truth segment are labeled the same ID as can be seen in Figure 3-10 (a). After that, surface growing segmentation is applied to process original data with ground truth IDs. From Figure 3-10(b), resulting segments with segment ID and ground truth ID will be obtained and the number of points in the biggest segment which overlaps ground truth segment are regarded as compared ones.

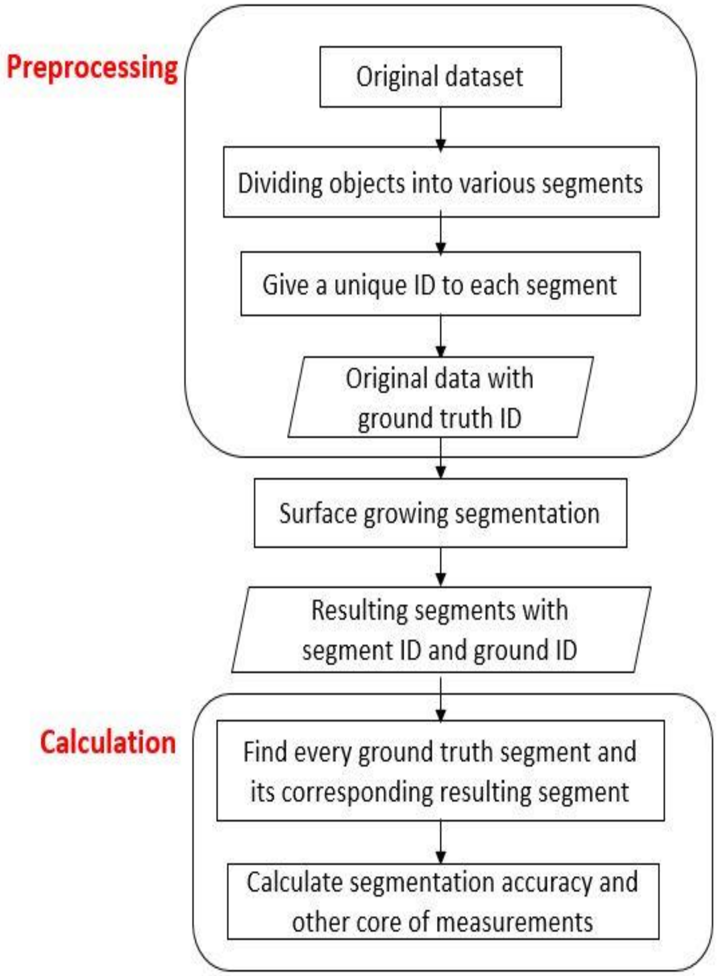
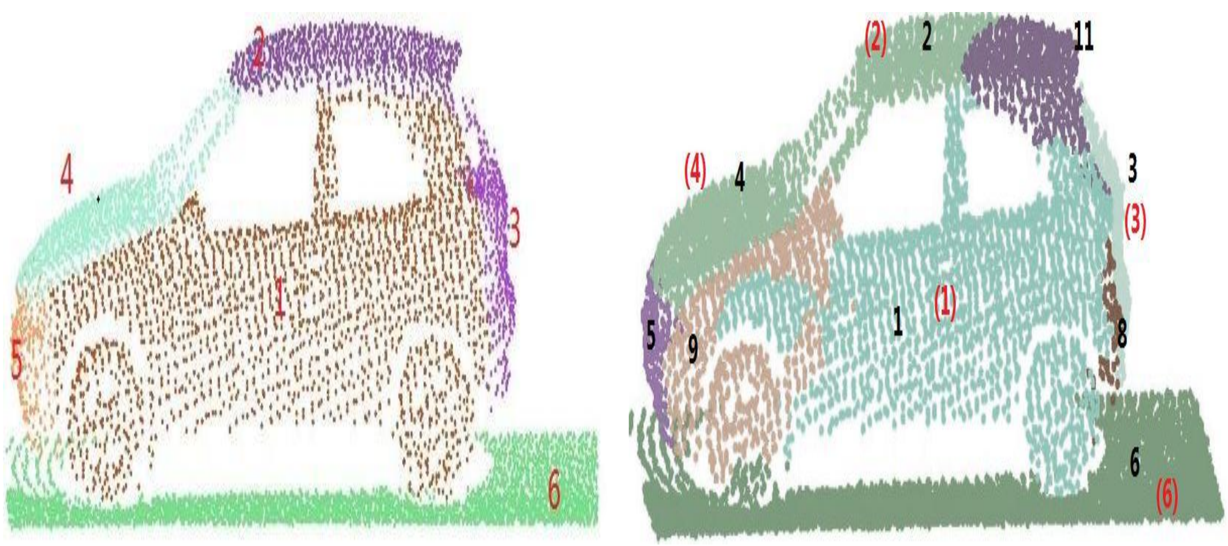


Figure 3-9 Block diagram of method. This method contain two stages, preprocessing and calculation



(a) Ground truth segments (b) Resulting segments

Figure 3-10 (a) Ground truth segments with ground truth IDs (red number) (b) Resulting segments with ground truth IDs (red number) and segment IDs (black number)

In the second stage, by quality assessment model, users can calculate the percentage of points in the biggest segments to that in ground truth segments., the percentage of points are under-segmented, the percentage of points are over-segmented, and the percentage of ground truth segments in all resulting segments. So four measures are calculated.

$$\text{AverAccuracy}=\frac{\text{The number of points within the corresponding biggest segments}}{\text{The total number of points}} \qquad 3\text{-}1$$

$$\text{OverSeg}=\frac{\text{The number of points wtihin other segments}}{\text{The total number of points}} \qquad 3\text{-}2$$

$$\text{UnderSeg}=\frac{\text{The number of points are assigned into the other biggest segments}}{\text{The total number of points}} \qquad 3\text{-}3$$

$$\text{GroundSeg}=\frac{\text{The number of groundtruth segments}}{\text{The total number of segments}} \qquad 3\text{-}4$$

Measure	**Optimum**
AverAccuracy	1 indicates perfect match
OverSeg	0.Minimum indicates perfect match or under-segmentation
UnderSeg	0.Minimum indicates perfect match or over-segmentation
GroundSeg	1 indicates perfect match

Table 3-1 Summary of the measures

3.3.2 Visual check

Visual assessment is just used for qualitative analysis. And it is much helpful to explain how over-segmentation occurs in this study.

3.4 Analyze effects of point distribution/point reduction on segmentation of MLS data

This part is divided into 2 phases. The first one is to analyze effects of point distribution on segmentation results. In this phase, find the optimal value of "Number of neighbors in kd-tree" in surface growing segmentation is a key issue, which can result in optimal segmentation results. In order to do this, a set of continuous values of "Number of neighbors in kd-tree" are applied to do segmentation, by quality assessment, select the best one as the optimal "Number of neighbors in kd-tree". Since surface growing is well-used to detect objects with planar segments, so segments in the wall are selected to visualize and calculate distribution of distance from every point to the fitted plane. Finally, influence of point distributions on segmentation of Paris and Enschede data are analyzed on the basis of visualization of resulting segments and its distribution of distance from points to the fitted plane.

The second one is to what impact of point reduction has on segmentation results. In order to make comparable with the segmentation accuracy of original data, other parameters in surface growing segmentation remain fixed, only change reduction parameter of every point reduction methods to get reduced datasets at different reduction levels. After that, calculate overall accuracy, every sample' accuracy like the first phase, then we can inspect segmentation accuracy is changed. In order to find out the reason why segmentation results of reduced data are different compared to that of original data. Only select segment in the wall to visualize and calculate the distribution of distance from every point to the fitted plane.

Finally, analysis on influence of point reduction on segmentation is done according to visualization of resulting segments in the wall and calculated distribution of distance from points to the fitted plane.

3.5 Evaluate applicability of every point reduction method

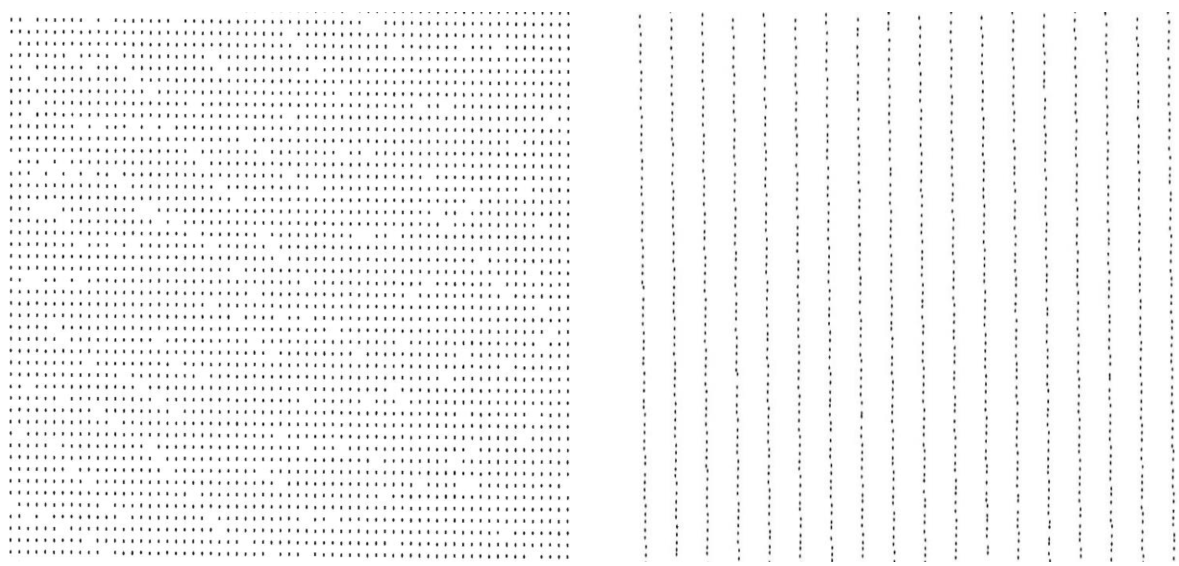
Applicability of every point reduction method on MLS dataset is evaluated according to the segmentation accuracy of direct point reduction results (facade), and iterative point reduction results (facade). Two measures are defined to assess every point reduction method' applicability: the range of segmentation accuracy and its levels of stability, user can determine which method to use while processing corresponding dataset.

4 Implementation and results

This chapter consist of introduction of study areas, the implementation of method, and results obtained.

4.1 Study area and dataset used

Paris and Enschede datasets are used in this study, which are acquired by mobile laser scanning with different configurations. Enschede dataset was acquired by OPTECH LYNX MOBILE MAPPER system with two laser scanners boarded on the car. And LYNX MOBILE MAPPER are installed non-orthogonally to the driving direction so as to get a complete point cloud of objects. The details of LYNX MOBILE MAPPER systems has been described in the review part. Figure 4-1 (a) shows Enschede data is quite even.



(a) Enschede data (b) Paris data

Figure 4-1 Visualization of Point distribution

Paris dataset is obtained by one laser scanner, which is like MDL DYNASCAN (Single MDL Laser scanner). Fig4-1(b) shows its point distribution, point density within every scanline is much denser than that in between neighboring scan lines.

Every dataset contains various kind of objects. Ground truth segments are selected manually in Paris and Enschede data. Caused by occlusion, some objects like bikes, pedestrians, even automobile interior decorations including seats, whose structure can't be described by enough points. It is a challenge to select ground truths of these objects, so these objects are not considered in this research. In order to calculate overall segmentation objectively and comprehensively. For every dataset, three different cars, traffic signs, ground samples, façade datasets, and trunks are selected, and these objects are selected in different areas of dataset. One group of ground truth segments in Paris dataset and Enschede data is shown in Figure 4-2 and Figure 4-3 separately.

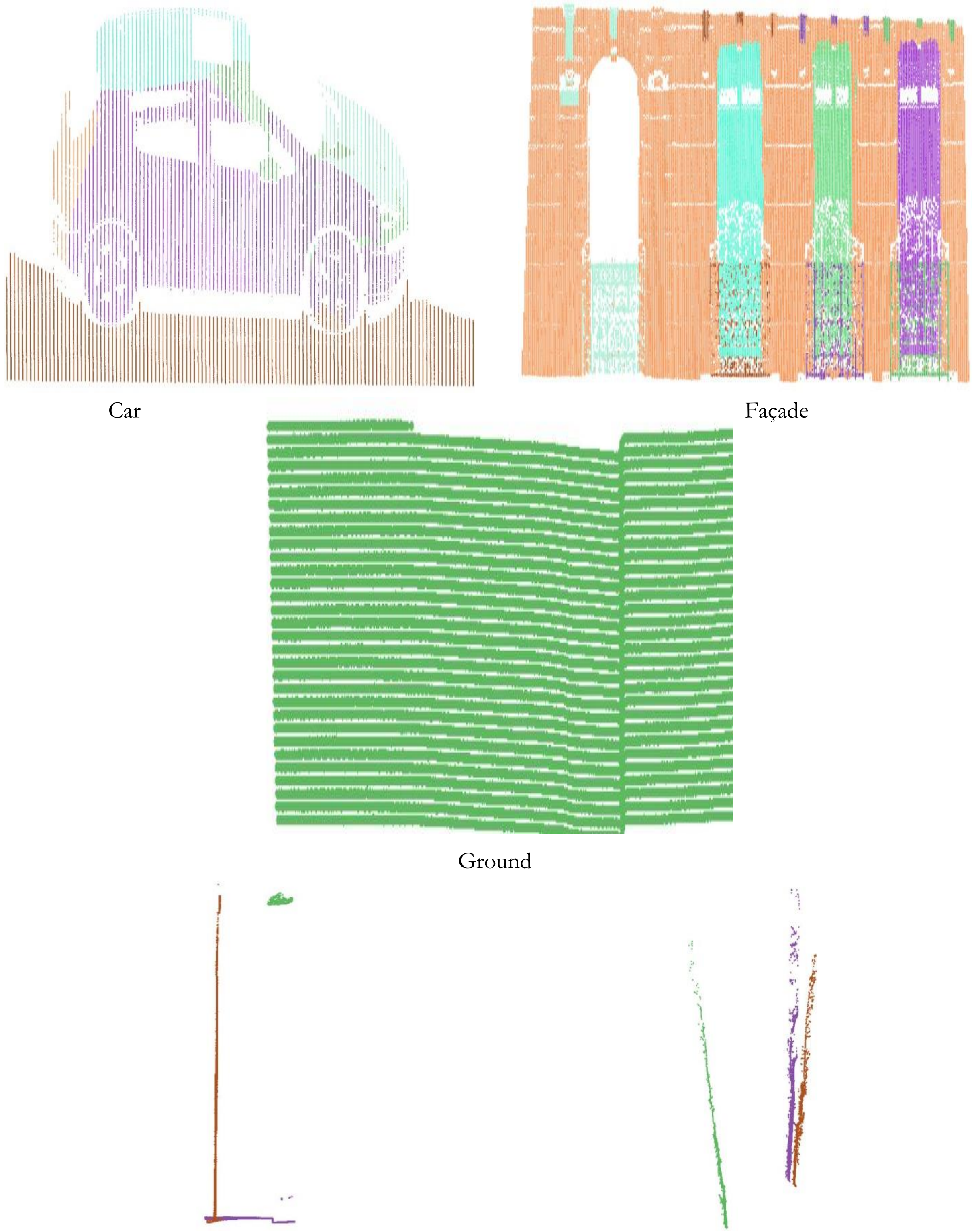


Figure 4-2 Ground truth segments in Paris dataset

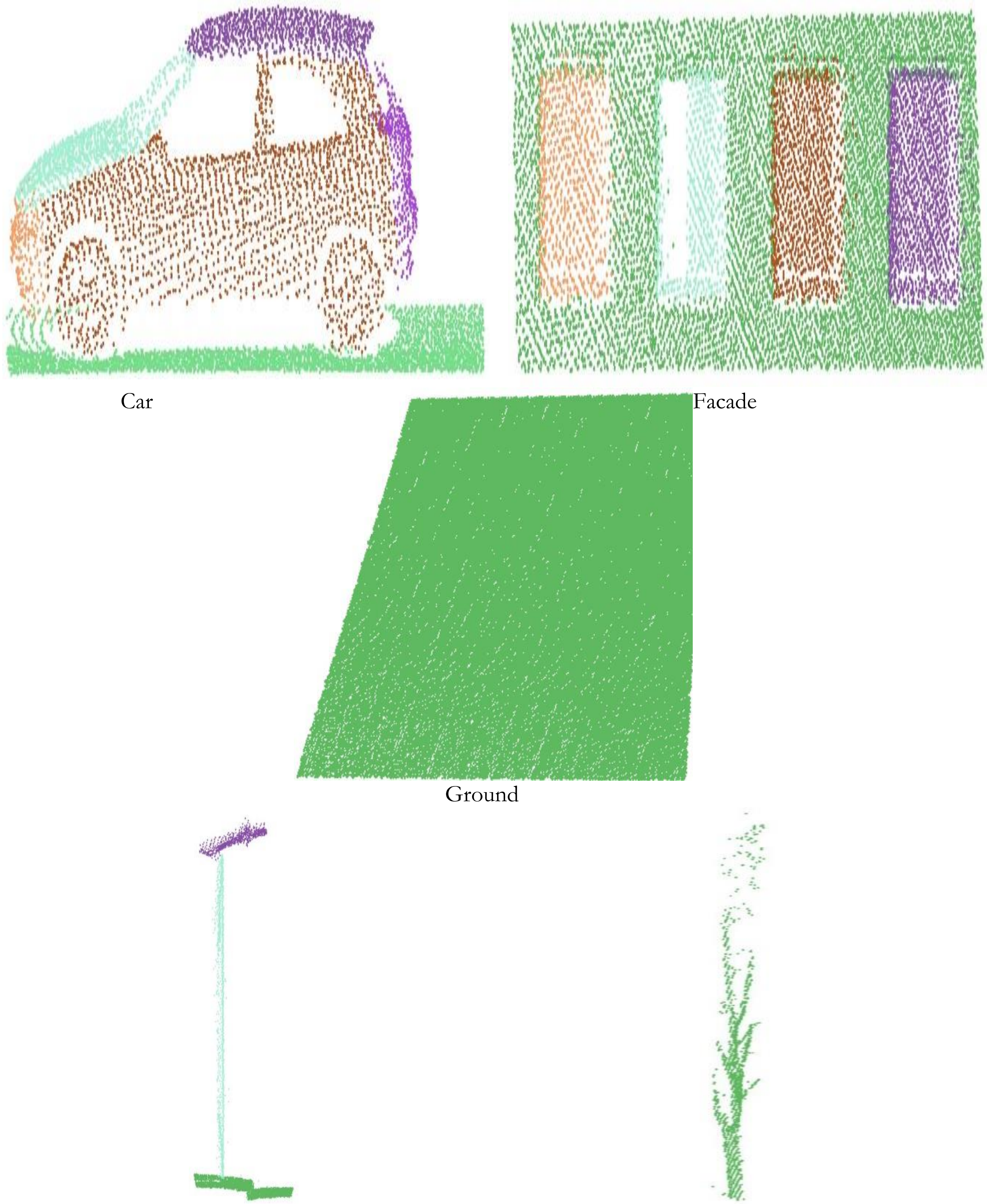
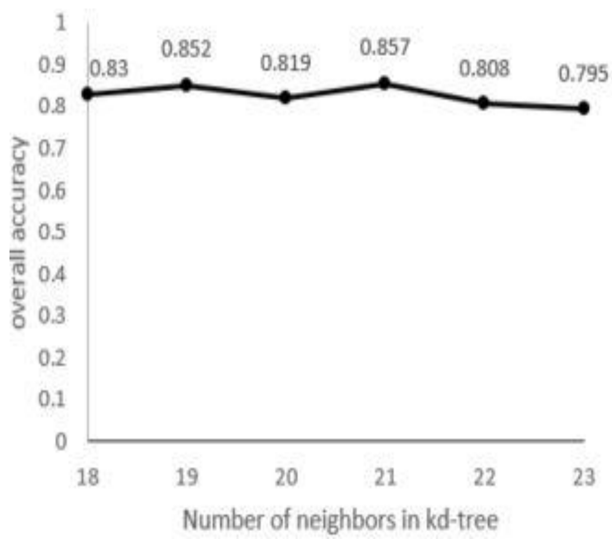


Figure 4-3 Ground truth segments in Enschede dataset

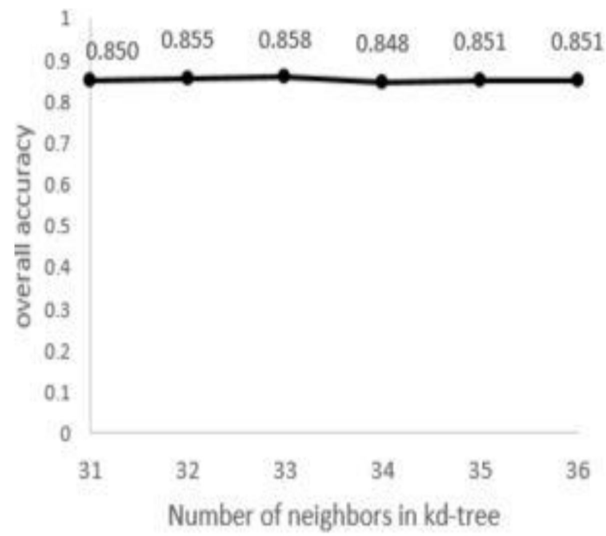
4.2 Analysis on effects of point distributions on segmentation results

Since point density of MLS data is very high, in surface growing segmentation, default value of "Number of neighbors in kd-tree" is set to 20 at first. Then in order to find the optimal segmenation results of point cloud, only fine tune value of "Number of neighbors in kd-tree", different ranges of this value are tried in this phase. Then select the optimal range and the best value of the range according to quality assessment results. Figure 4-4 (a) shows the highest segmentation accuracy of Paris data is 0.857, the related "the number of neighbors in kd-tree" is 21. Figure 4-4 (b) shows the highest segmentation accuracy of Enschede data is 0.858, the related "the number of neighbors in kd-tree" is 33. Values of core measures calculated based on Paris data and Enschede data are shown Table 4-1 and Table 4-2. Only one group of segmentation results from Paris data and Enschede data is shown in Figure 4-5 and Figure 4-6. Since there is no error with ground segmentation, there is no necessity to show its results. Even through there is no significant differences between these two values, accuracy of façade data set is much different. One façade dataset is selected to calculate its values of core measures so as to analyze what the impact of point distribution has on segmentation results. As shown in Table 4-3, segmentation accuracy of façade in Paris data is much lower than that of Enschede data, because there is too much over-segmentation of façade in Paris data.

One façade dataset is selected to calculate its values of core measures so as to analyze what the impact of point distribution has on segmentation results. As shown in Table 4-3, segmentation accuracy of façade in Paris data is also much lower than that of Enschede data, because there is much over-segmentation of façade in Paris data. In order to investigate why it occurred. Only select segments in the wall and calculate the distribution of distance from every point to the fitted plane, which are shown in Figure 4-7.



(a) Paris data



(b) Enschede data

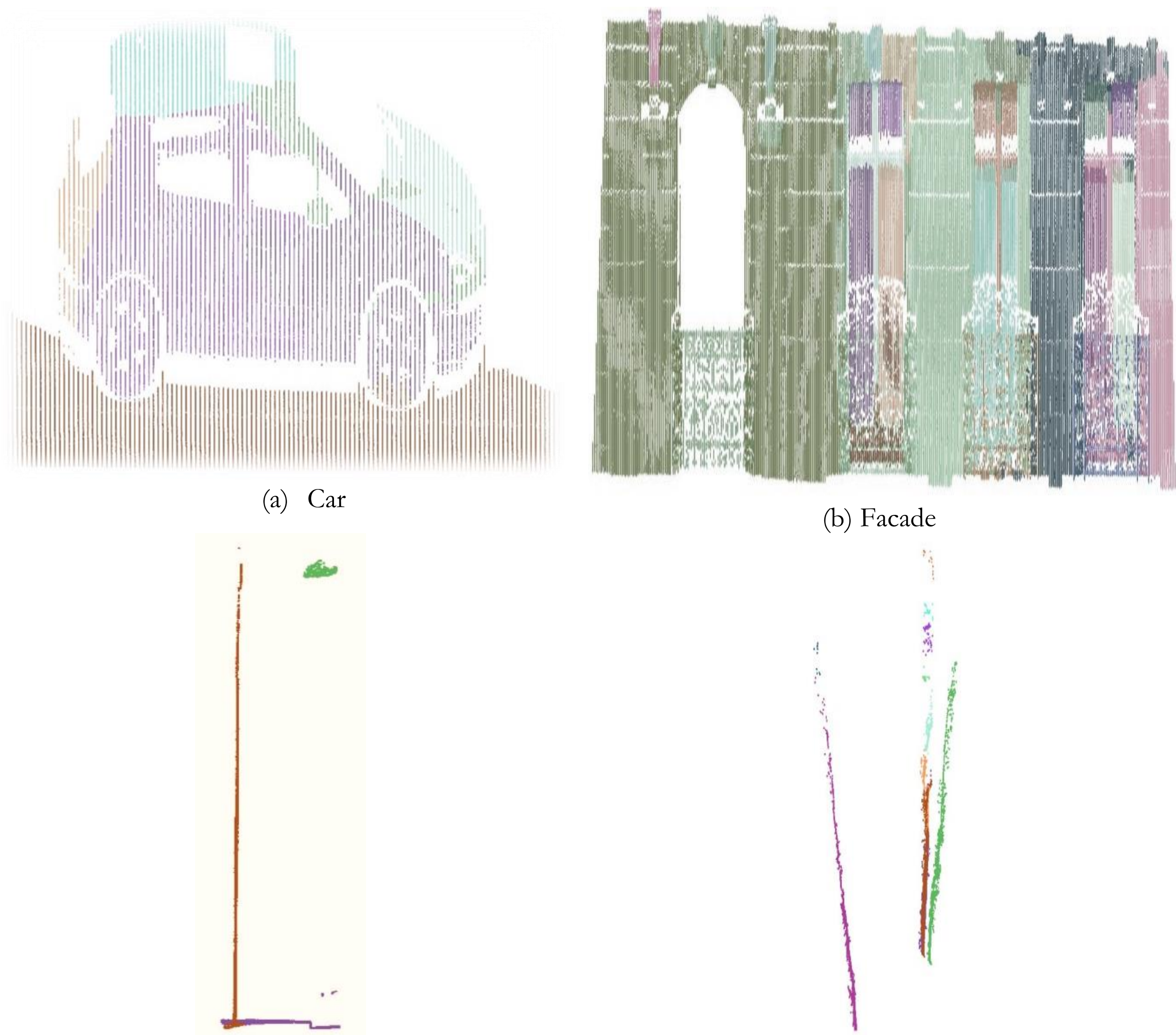
Figure 4-4 Optimal segmentation results selection

Segments in Objects	Num Points	Aver Accuracy	OverSeg	UnderSeg	GroundSeg
Car	29773	0.873	0.049	0.078	0.606
Façade	48014	0.842	0.134	0.024	0.372
Ground	39789	1.000	0	0	1.000
Sign	4539	0.945	0.022	0.034	0.850
Trunk	5922	0.685	0.297	0	0.242
Average	28870	0.857	0.102	0.042	0.545

Table 4-1 Values of core measures (Paris data)

Segments in Objects	Num Points	Aver Accuracy	OverSeg	UnderSeg	GroundSeg
Car	21260	0.835	0.15	0.015	0.717
Façade	9914	0.924	0.067	0.008	0.476
Ground	67507	1	0	0	1.000
Sign	14202	0.939	0.059	0.003	0.65
Trunk	1088	0.454	0.546	0	0.126
Average	19642	0.858	0.133	0.009	0.624

Table 4-2 Values of core measures (Enschede data)



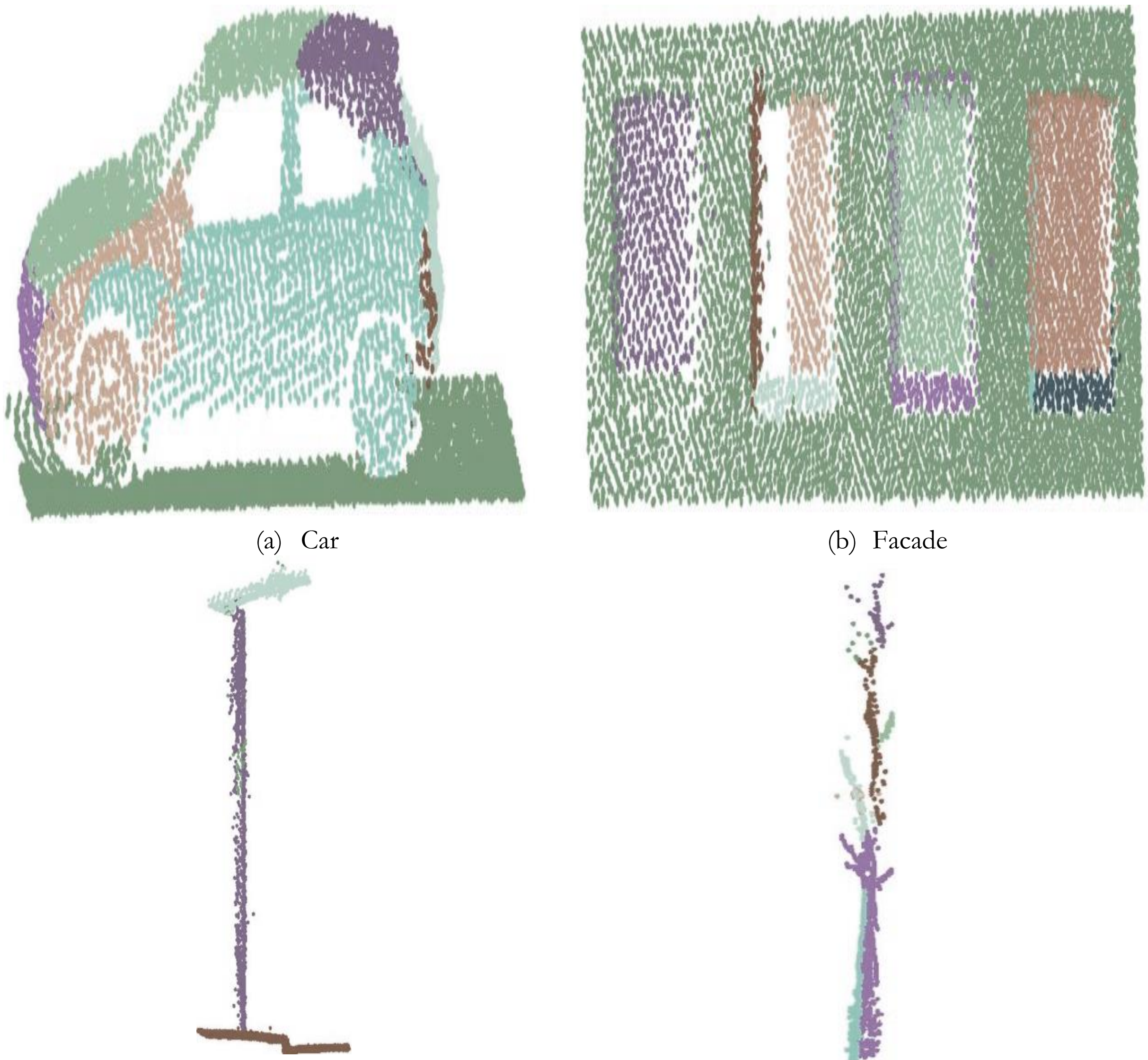
(a) Car

(b) Facade

(c) Sign

(d) Trunk

Figure 4-5 Segmentation results (Paris data)



(a) Car

(b) Facade

(c) Sign

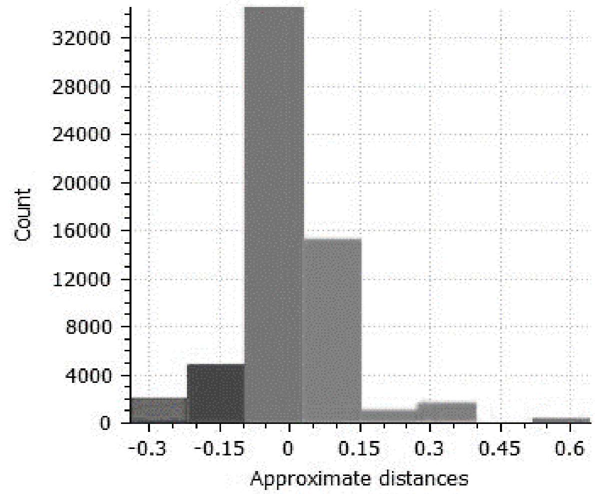
(d) Trunk

Figure 4-6 Segmentation results (Enschede data)

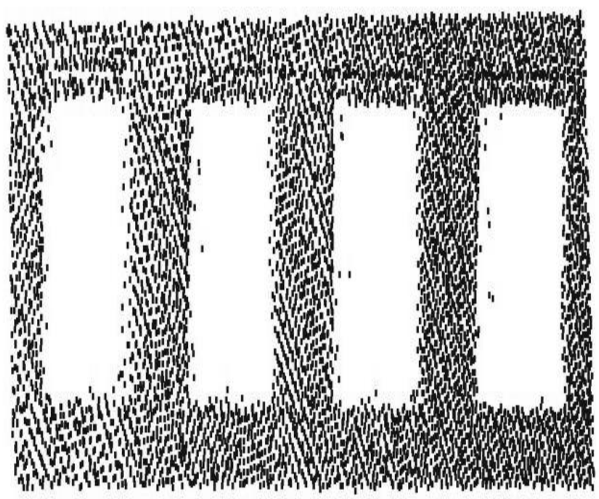
	accuracy	OverSeg	UnderSeg	GroundSeg
(a)	0.51	0.316	0.174	0.32
(b)	0.914	0.075	0.011	0.333

Table 4-3 (a) Values of core measures (façade in Paris data) (b) Values of core measure (façade in Enschede data)

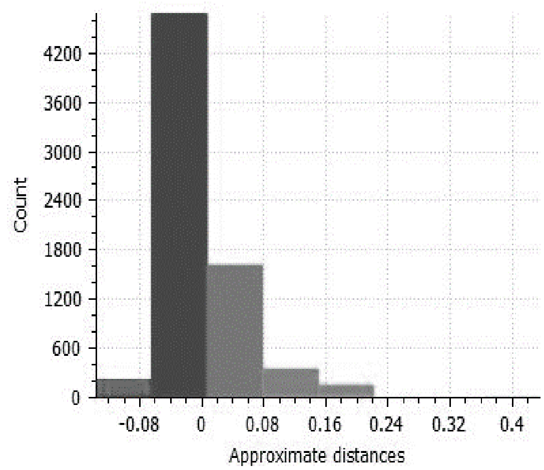
(a) Segments in the wall (Paris data)



(b) Distribution of distance from every point to the fitted plane(Paris data)



(c) Segments in the wall (Enschede data)



(a) Distribution of distance from every point to the fitted plane (Enschede data)

Figure 4-7 Visualization of point distribution on the wall

As shown in Figure 4-7 (a), in the segments of wall, there is much over-segmentation. Since "Maximum distance to the surface" of surface growing is set to 0.2. From Figure 4-8 (b), most of points fall into the range between [-0.2,0.2], a tiny part of points are over this range. Generally, most points should be in one segment, only a few points are assigned into two or more segments. Results show most points are assigned into two or more segments because of the jumping part on the wall. The jumping part is shown in Figure 4-8.

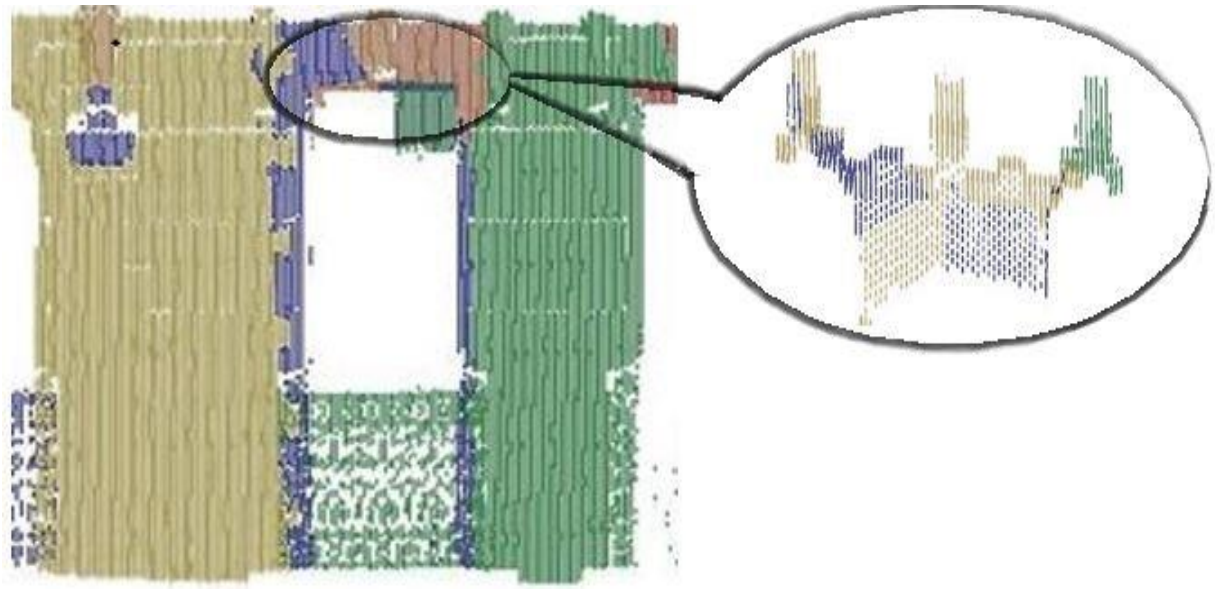
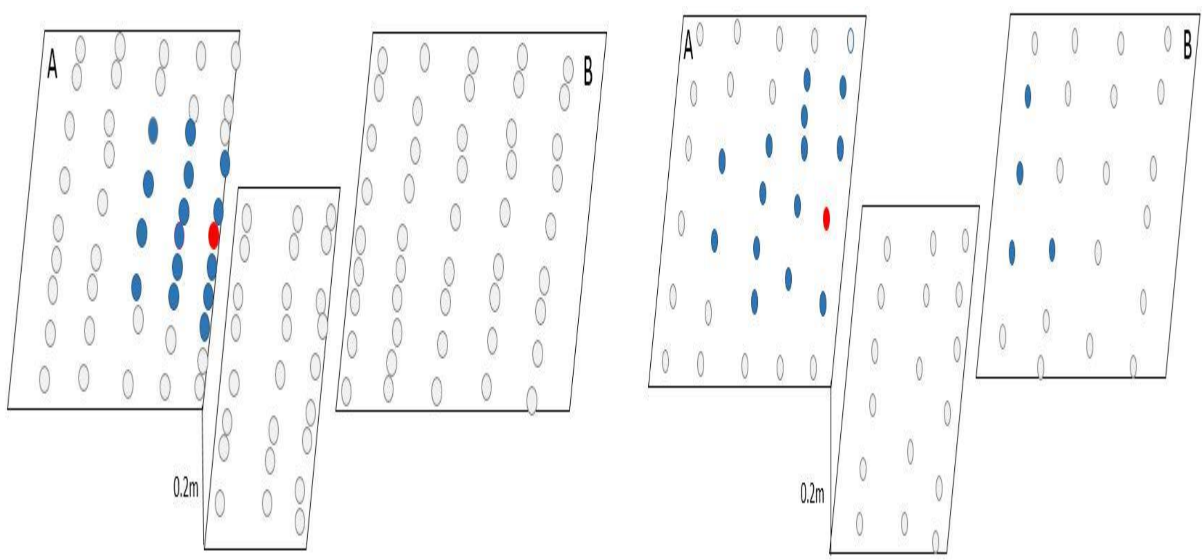


Figure 4-8 Visualization of the jumping part between segments on the wall (Top view)

From Figure 4-9,when these flat parts want to grow into one segment, the seed will search the nearest 21 points within 1m around it. Then distance from every candidate point to the fitted plane is calculated to see whether it is less than 0.2m. If it is, the candidate will be accepted, vice versa. Since point density of Paris is uneven, point density of points in one scan line is much higher than that between neighboring scan lines. So the red seed point in plane A can not find any candidate point in plane B. So these flat parts are separted by the jumping part.



(a) Point distribution (Paris dataset)　　(b) Point distribution (Enschede dataset)

Figure 4-9 Visualization of point distribution

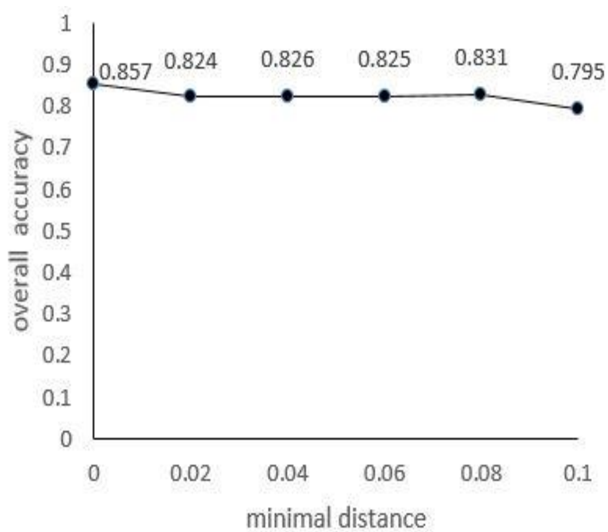
On the contrary, there is no over-segmentation in the wall in Enschede datasret. One of reasons is that structure of façade is much regular. The other one is that Enschede data is much even which is shown in Figrue 4-9(b). The red seed point in plane A can find suitable candidate points from plane B, so these flat parts can be merged into one segment.

4.3 Analysis on effects of point reduction on segmentation of Paris data

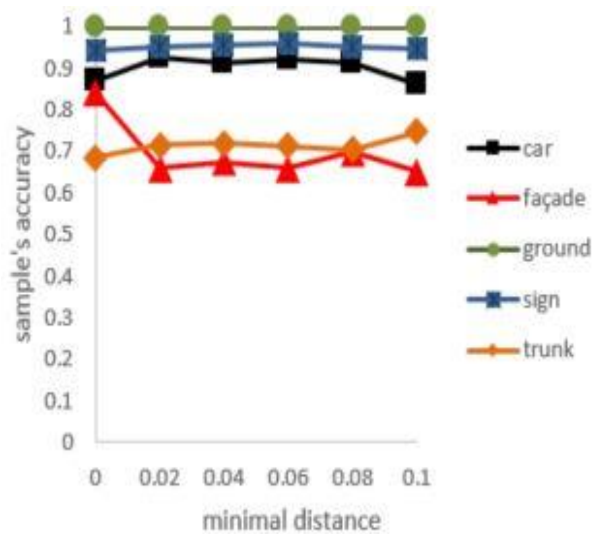
Six existing point reduction methods mentioned in section 3.2 are used to reduce Paris dataset. In order to analyze what impact of these methods have on segmentation of Paris dataset, parameters of surface growing segmentation used for processing all reduced Paris datasets remain the same, specially, the "number of neighbors in kd-tree" is set to 21. Only change reduction parameters of every point reduction method to get resulting datasets at different reduction levels, which is called "direct point reduction". Then calculate overall accuracy, every sample data's accuracy, and select one façade dataset to show its accuracy at different point reduction levels. There is another case that users may do point reduction based on original-reduced dataset over and over again, which is called "iterative point reduction". For the time being, one façade original-reduced dataset with the worst segmentation accuracy is selected to do iterative point reduction to see whether it become worse or better. Finally, select segments within wall from original-reduced and original-reduced-reduced dataset to compute the distribution of distances from every point to the fitted plane so as to analyze influence of point reduction on segmentation results.

For method 1:

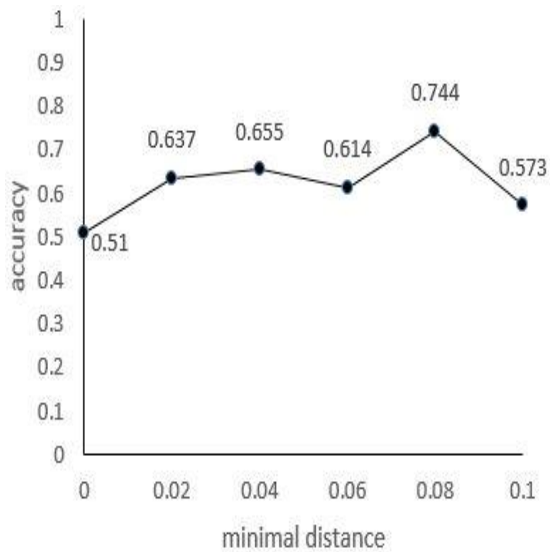
From Table 4-4, value of core measures are a little different between original-reduced dataset and original-reduced-reduced dataset, but from Figure 4-11, there is no change of segmentation results between these two datasets. Since method 1 is to remove points on the basis of distance, the distance is set to 0.1m, so after direct point reduction once, the distance among neighboring points is 0.1m, then regardless of how many times running this method with reduction parameter of "0.1m" to process data, the distance among neighboring points always remain the same, no points are deleted.



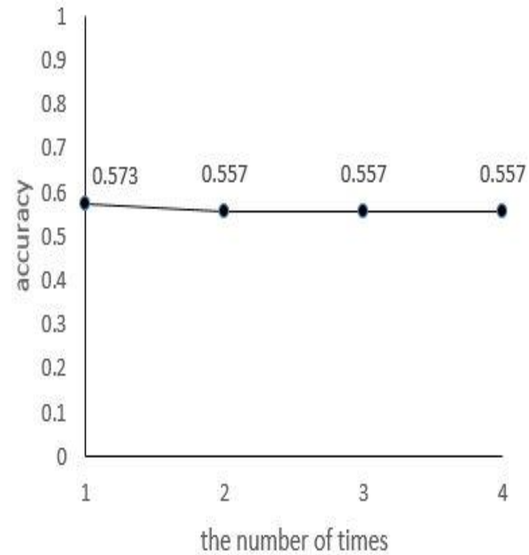
(a) Overall segmentation accuracy



(b) Every sample's segmentation accuracy



(c) Segmentation accuracy of direct point reduction dataset(one façade dataset)

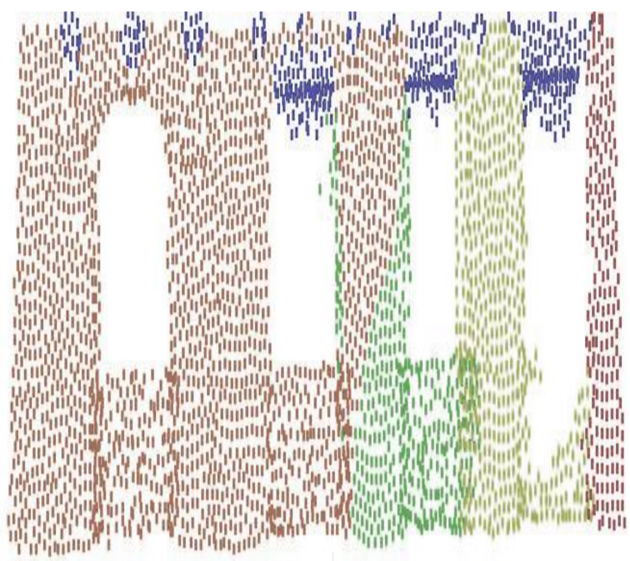


(d) Segmentation accuracy of iterative point reduction dataset(one façade dataset)

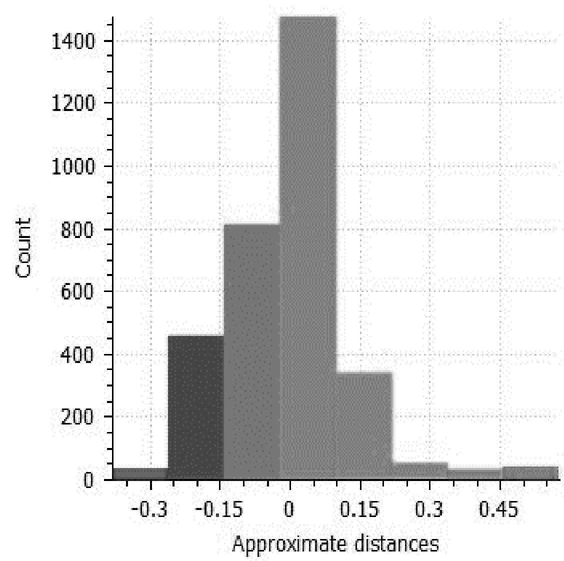
Figure 4-10 Segmentation accuracy

	Reduction Times	Reduction parameter	Remaining rate	Accuracy	OverSeg	UnderSeg	GroundSeg
(a)	1	0.1m	5.8%	0.573	0.279	0.148	0.571
(b)	2	0.1m	5.5%	0.557	0.312	0.132	0.5

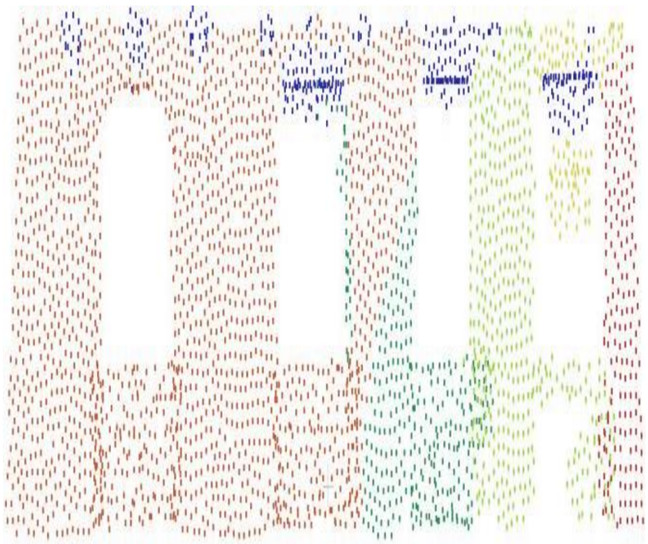
Table 4-4 (a) Values of core measures from original-reduced data (façade) (b) Values of core measures based on original-reduced-reduced data (facade)



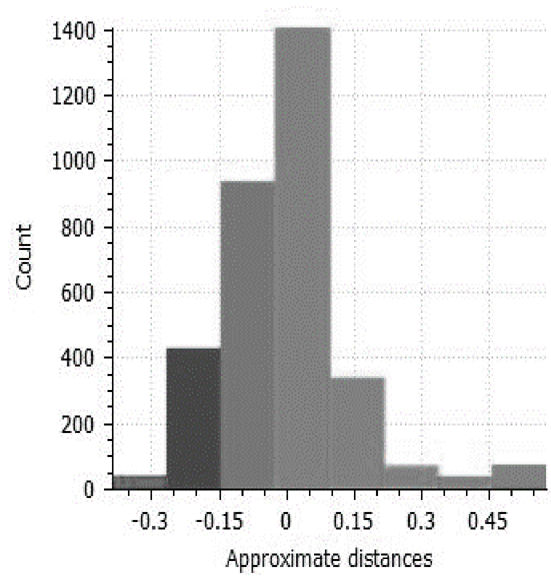
(a) Segments within wall based on original-reduced dataset



(b) Distances from points to the fitted plane (original-reduced dataset)



(c) Segments within wall based on original-reduced-reduced dataset

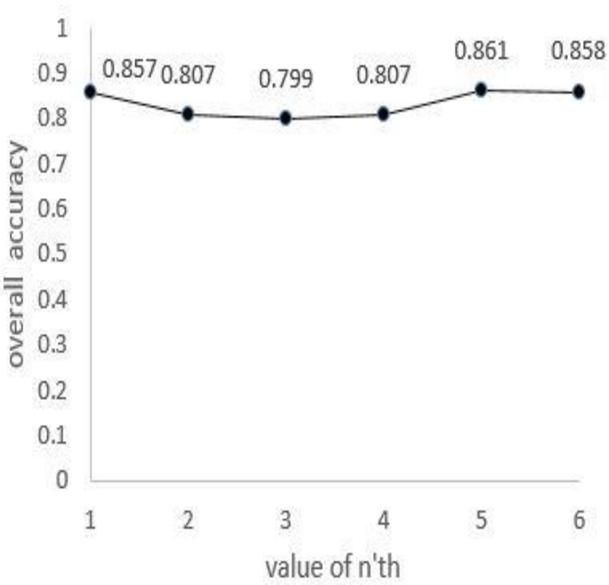


(d) Distances from points to the fitted plane(original-reduced-reduced dataset)

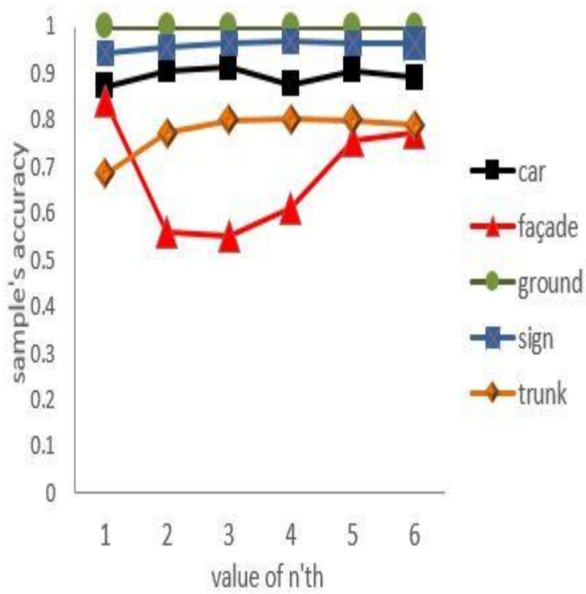
Figure 4-11 Visualization of point distribution on the wall

For method 2:

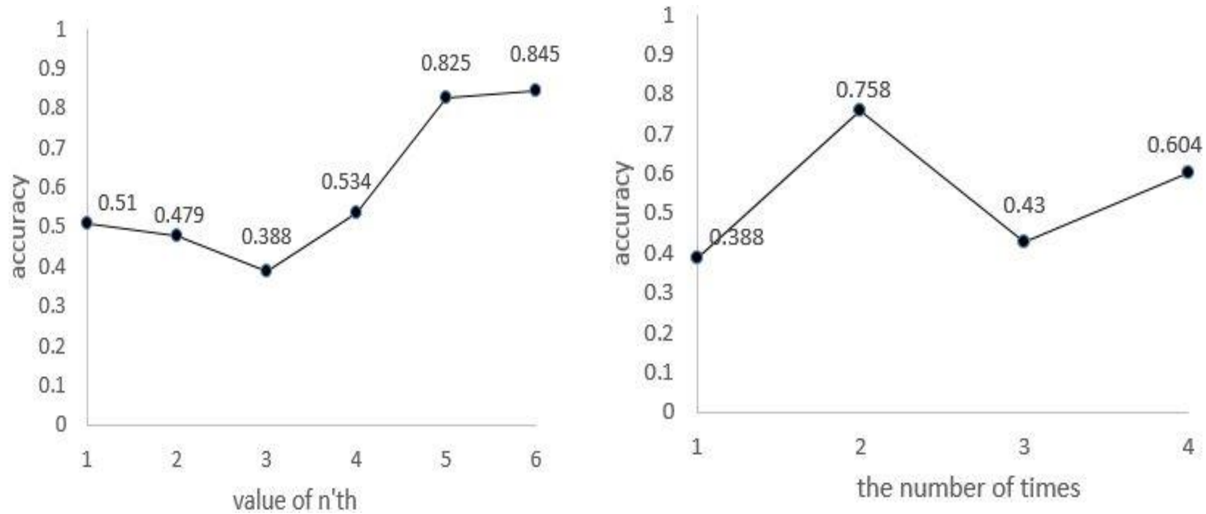
From Table 4-5 and Figure 4-13, at the same point reduction level, direct point reduction and iterative point reduction can lead to the same segmentation result. So at the same remaining rate, regardless of segmentation results obtained by direct point reduction or iterative point reduction, resulting point distribution remain the same.



(a) Overall segmentation accuracy



(b) Every sample's segmentation accuracy



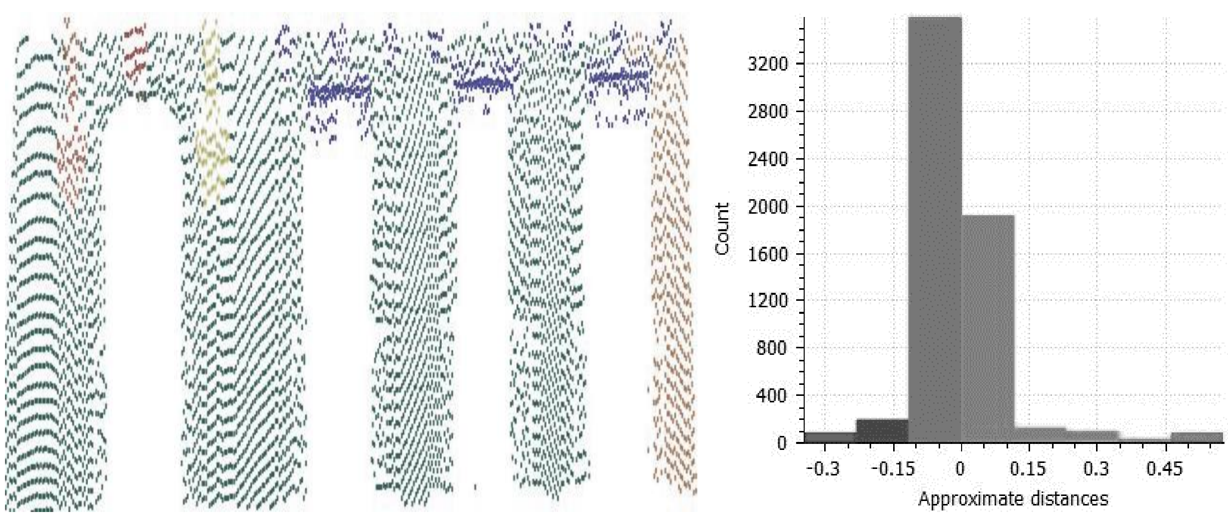
(c) Segmentation accuracy of direct point reduction dataset (one façade dataset)

(d) Segmentation accuracy of iterative point reduction dataset(one façade dataset)

Figure 4-12 Segmentation accuracy

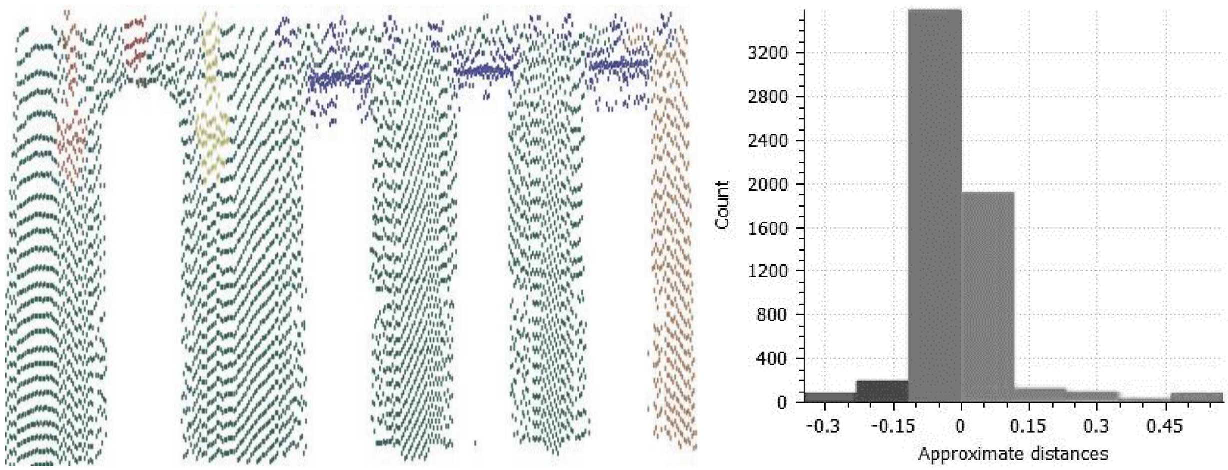
	Reduction Times	Reduction parameter	Remaining rate	Accuracy	OverSeg	UnderSeg	GroundSeg
(a)	1	9	11.2%	0.758	0.228	0.014	0.4
(b)	2	3	11.2%	0.758	0.228	0.014	0.4

Table 4-5 (a) Values of core measures from original-reduced data (façade) (b) Values of core measures based on original-reduced-reduced data (facade)



(a) Segments within wall based on original-reduced dataset

(b) Distances from points to the fitted plane(original-reduced dataset)



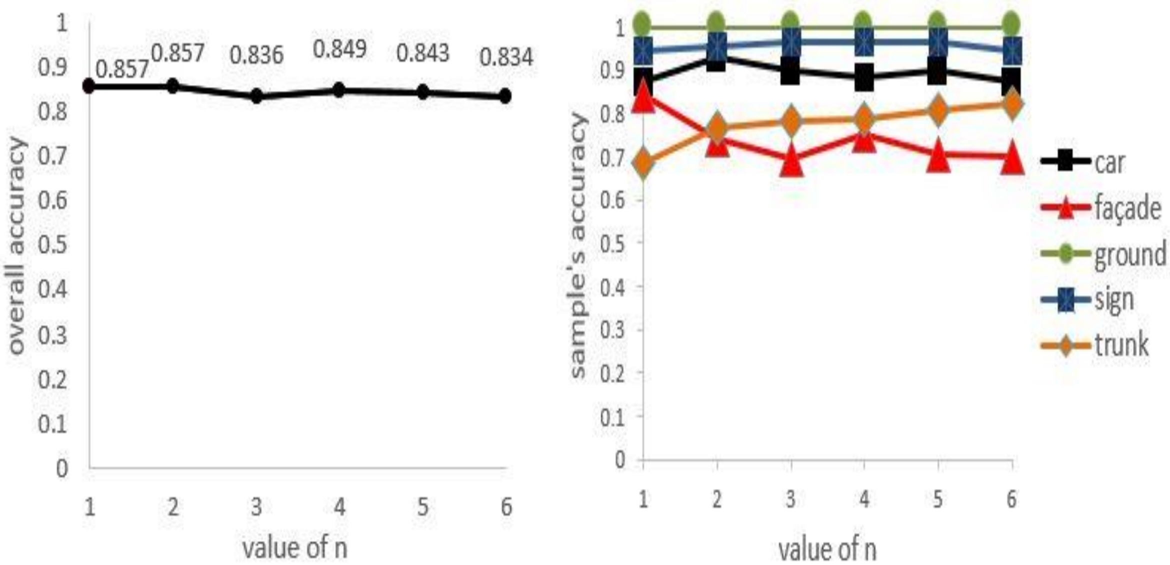
(c) Segments within wall based on original-reduced-reduced dataset

(d) Distances from points to the fitted plane (original-reduced-reduced dataset)

Figure 4-13 Visualization of point distribution on the wall

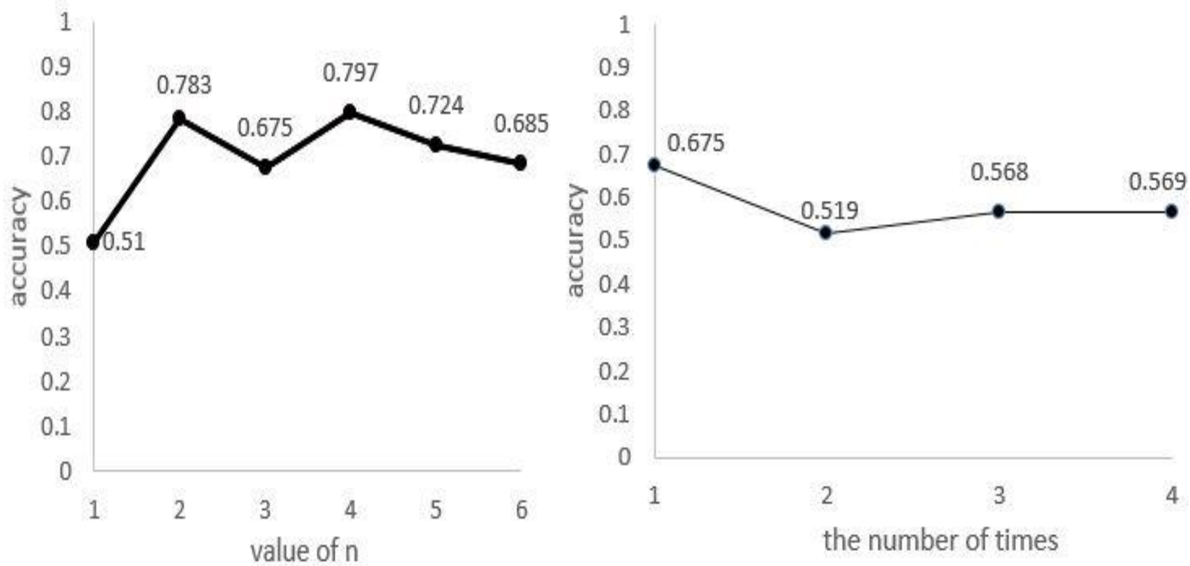
For method 3:

From Figure 4-14(d), after iterative point reduction, segmentation accuracy of point is decreasing. And From Table 4-6, at same remaining rate, quality of segmentation obtained by direct point reduction is much higher than that obtained by iterative point reduction. The segmentation results are shown in Figure 4-15. However, this method has bad Stability, which is to remove point cloud by randomly selecting one point within every set of n points. So these results are based on one group of resulting point distributions.



(a) Overall segmentation accuracy

(b) Every sample's segmentation accuracy



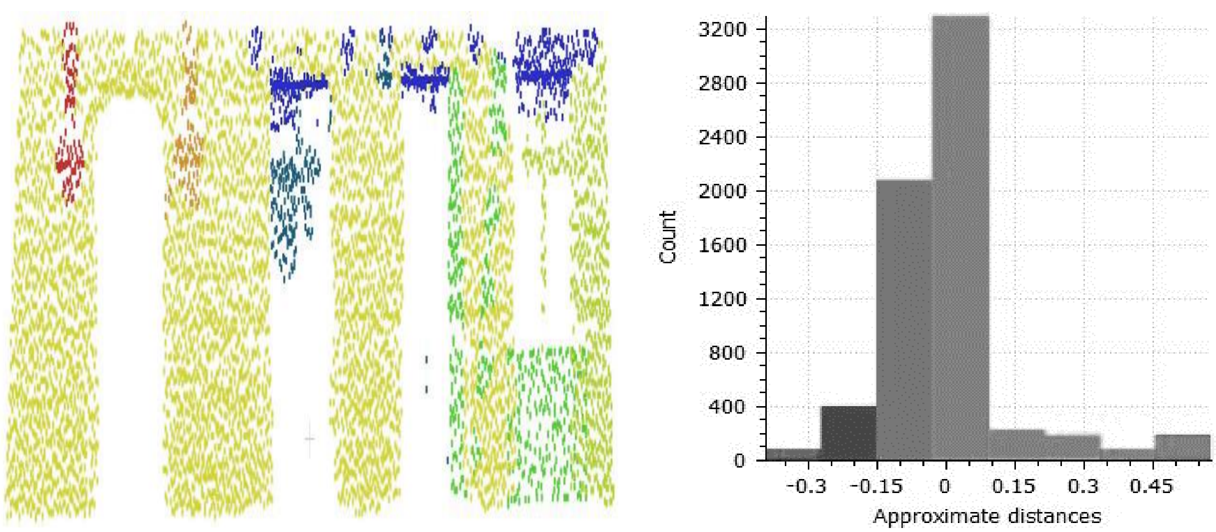
(c) Segmentation accuracy of direct point reduction dataset(one façade dataset)

(d) Segmentation accuracy of iterative point reduction dataset (one of façade datasets)

Figure 4-14 Segmentation accuracy

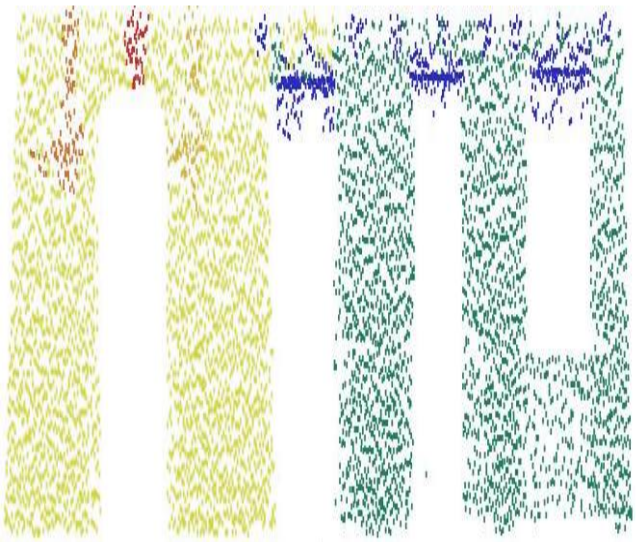
	Reduction Times	Reduction parameter	Remaining rate	Accuracy	OverSeg	UnderSeg	GroundSeg
(a)	1	9	11.2%	0.728	0.233	0.039	0.444
(b)	2	3	11.2%	0.519	0.189	0.292	0.421

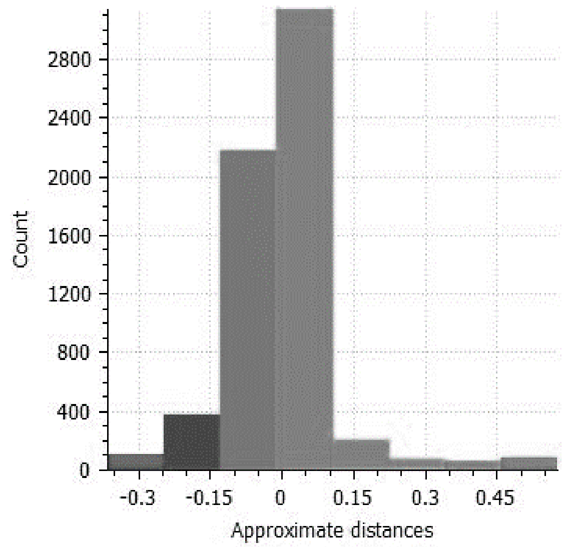
Table 4-6 (a) Values of core measures from original-reduced data (façade) (b) Values of core measures based on original-reduced-reduced data (facade)



(a) Segments within wall based on original-reduced dataset

(b) Distances from points to the fitted plane(original-reduced dataset)





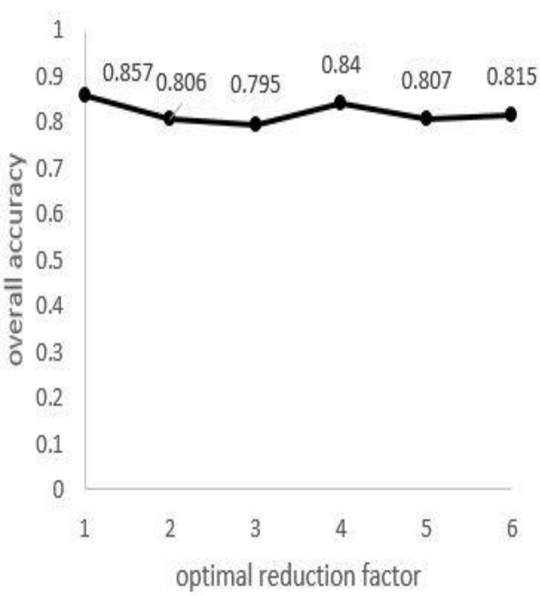
(c) Segments within wall based on original-reduced-reduced dataset

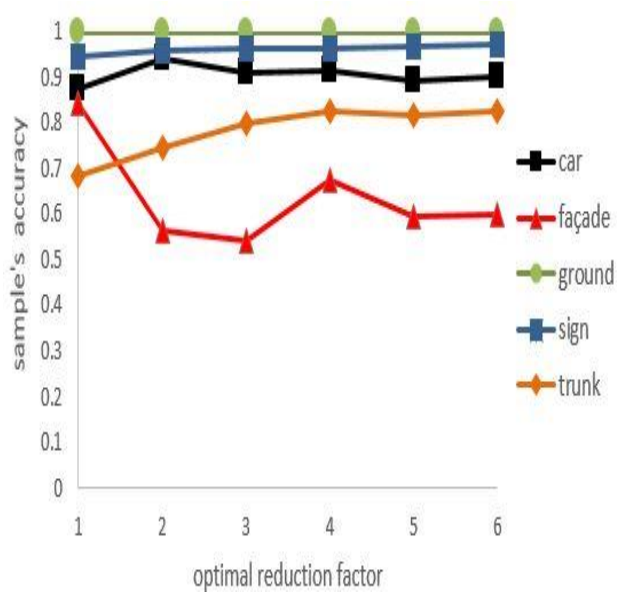
(d) Distances from points to the fitted plane(original-reduced-reduced dataset)

Figure 4-15 Visualization of point distribution on the wall

For method 4:

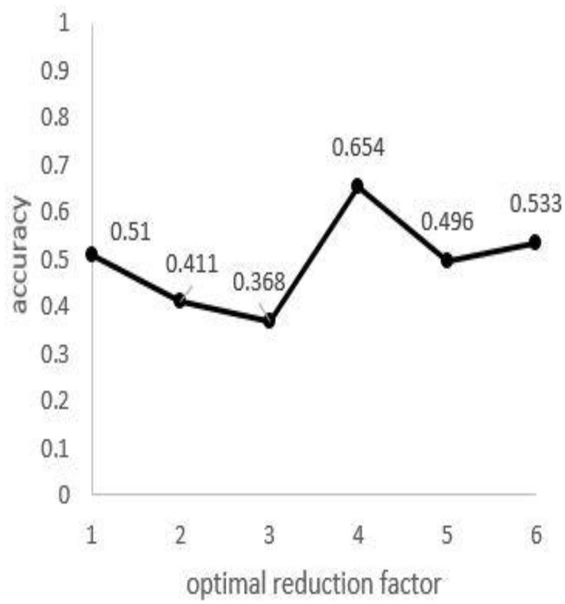
From Figure 4-17 (d), after iterative point reduction, segmentation accuracy is increasing. And From Table 4-7, at the same remaining rate, quality of segmentation obtained by iterative point reduction is just a little higher than that obtained by direct point reduction. The related point distributions are shown in Figure 4-18. Like method 3, this method is also instable, since the selection of seed points is random. Even though every time applying this method with the same reduction parameter to reduce the same dataset, the resulting point distributions are not the same. So these results are just based on one set of reduced dataset.



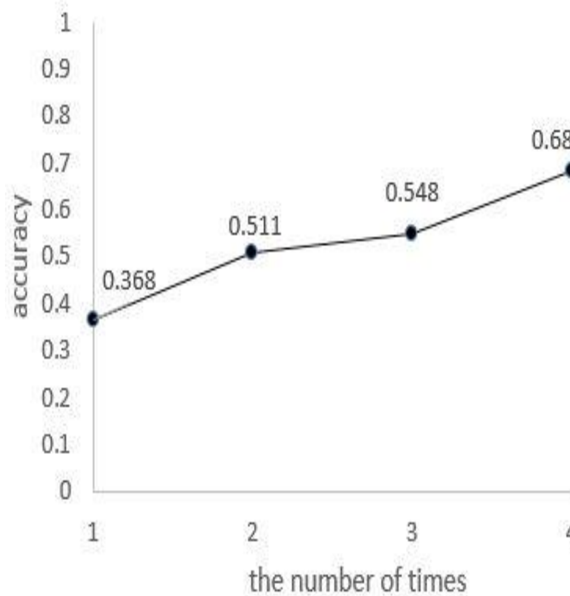


(a) Overall segmentation accuracy

(b) Every sample's segmentation accuracy



(c) Segmentation accuracy of direct point reduction dataset(one façade dataset)

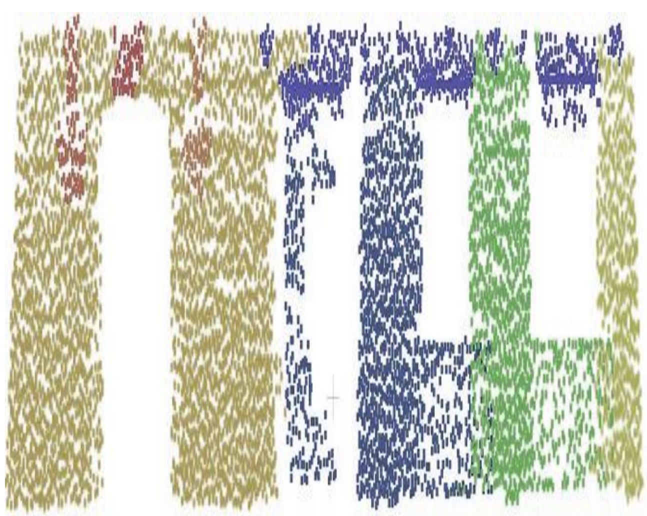


(d) Segmentation accuracy of iterative point reduction dataset (one façade dataset)

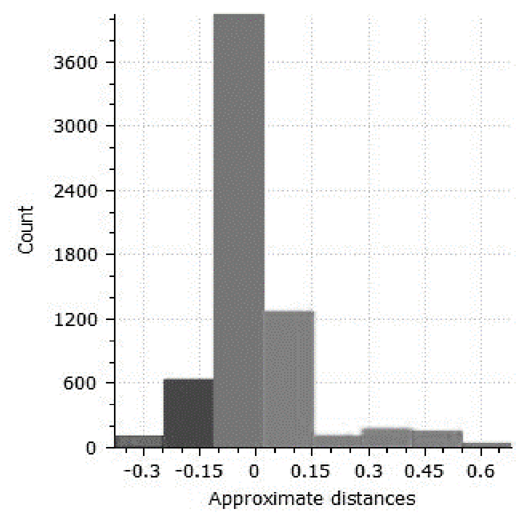
Figure 4-16 Segmentation accuracy

	Reduction Times	Reduction parameter	Remaining rate	Accuracy	OverSeg	UnderSeg	GroundSeg
(a)	1	9	11.2%	0.51	0.256	0.235	0.421
(b)	2	3	11.2%	0.511	0.379	0.109	0.381

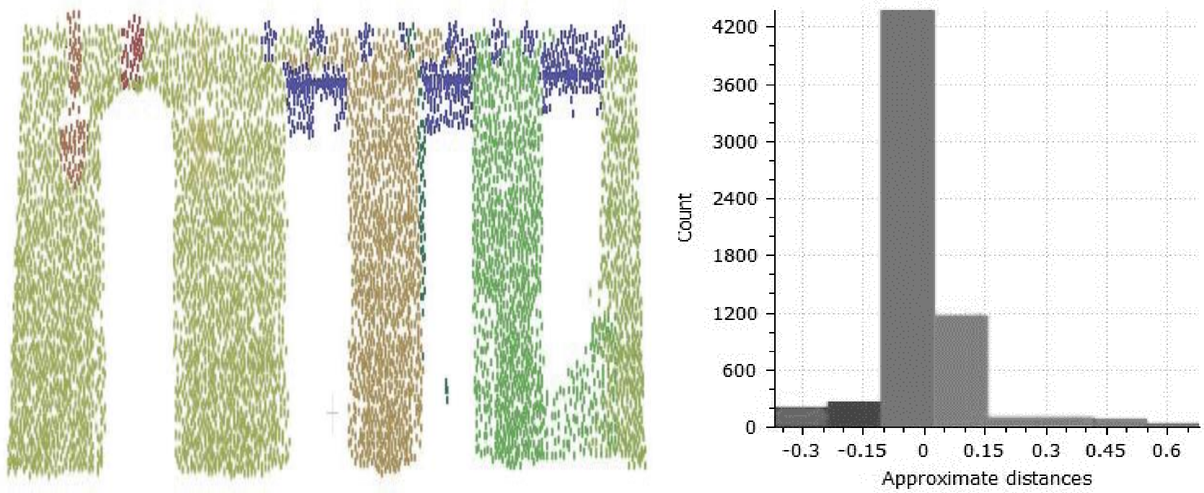
Table 4-7 (a) Values of core measures from original-reduced data (façade) (b) Values of core measures based on original-reduced-reduced data (façade)



(a) Segments within wall based on original-reduced dataset



(b) Distances from points to the fitted plane (original-reduced dataset)



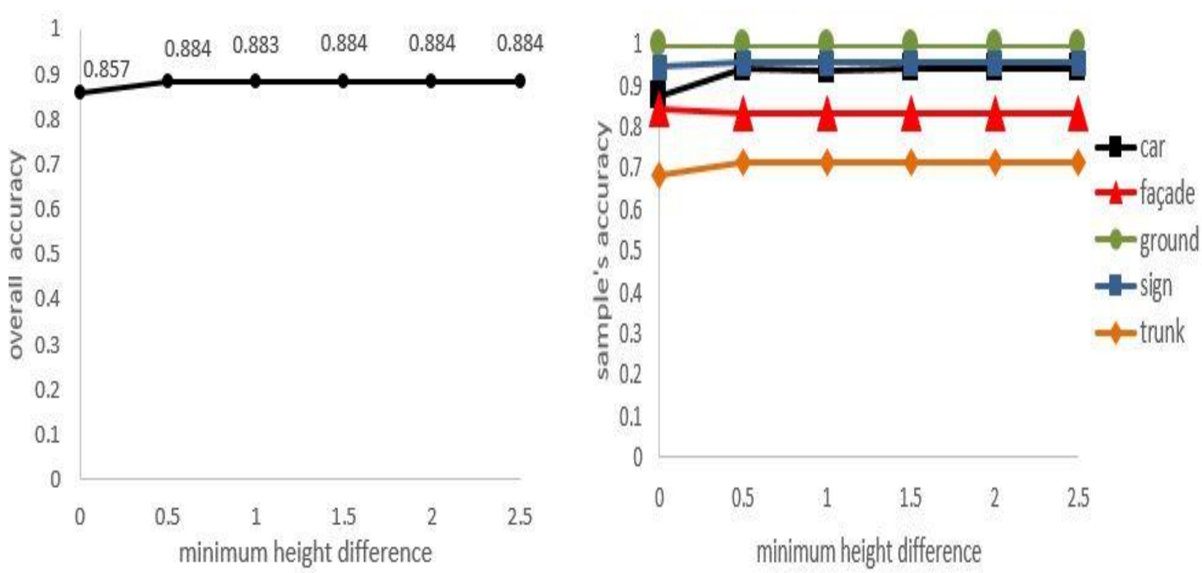
(c) Segments within wall based on original-reduced-reduced dataset

(d) Distances from points to the fitted plane (original-reduced-reduced)

Figure 4-17 Visualization of point distribution on the wall

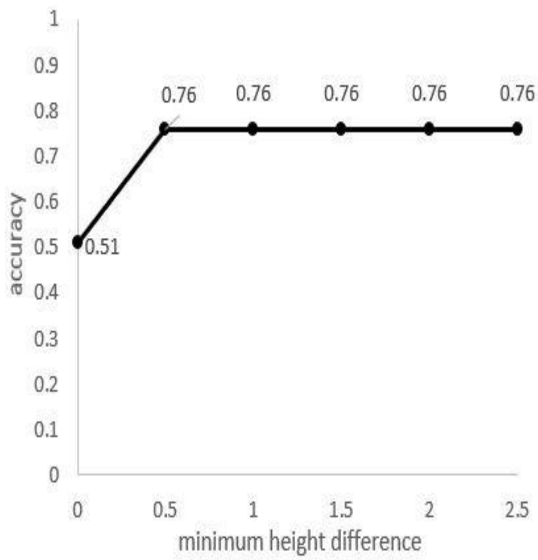
For method 5:

Since scanning frequency of laser scanning device is very high, which always result in over-resample. So some points has same X, Y coordinates, this method is to remove one of two points with the same X, Y coordinates. From Figure 4-18 (c), in the direct point reduction, no matter what the point reduction parameters set in this method, only one of two points with same X, Y is removed. From Figure 4-18 (d), in the iterative point reduction, regardless of how many times this method are used, still only one of two point with same X, Y coordinates are removed, there are no points with the same X, Y coordinates since point reduction twice, so segmentation accuracy remain stable afterwards. From Table 4-8, regardless of which point reduction way to be used to reduce dataset, remaining rate is still high. In addition, in iterative point reduction, segmentation accuracy is worse than that of original data. The reason is the first time deleted points contain double points. The second time deleted point has non-double points. Overall, method 5 is not well-used for Paris dataset.

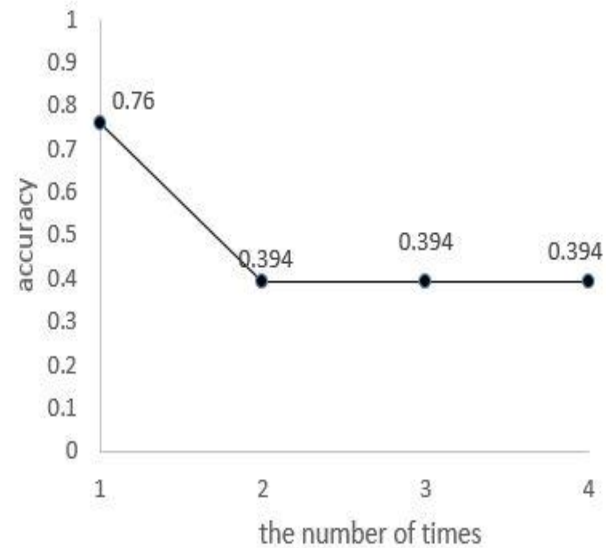


(a) Overall segmentation accuracy

(b) Every sample's segmentation accuracy



(c)Segmentation accuracy of direct point reduction dataset (one façade dataset)



(d) Segmentation accuracy of iterative point reduction dataset dataset (one façade dataset)

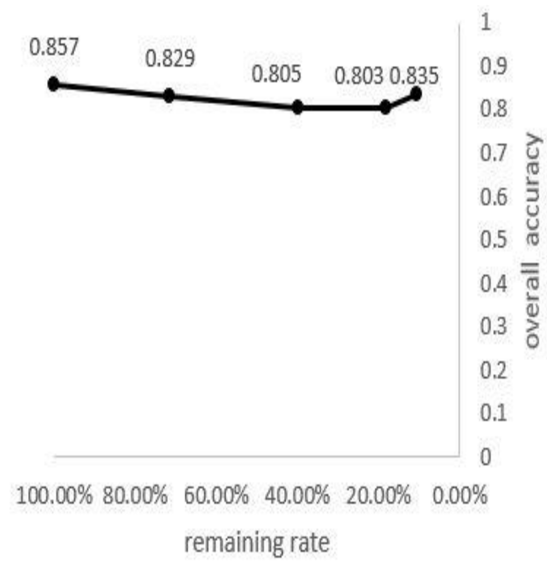
Figure 4-18 Segmentation accuracy

	Reduction Times	Remaining rate	accuracy	OverSeg	UnderSeg	GroundSeg
(a)	1	89.1%	0.76	0.197	0.043	0.242
(b)	2	60.1%	0.394	0.397	0.209	0.32

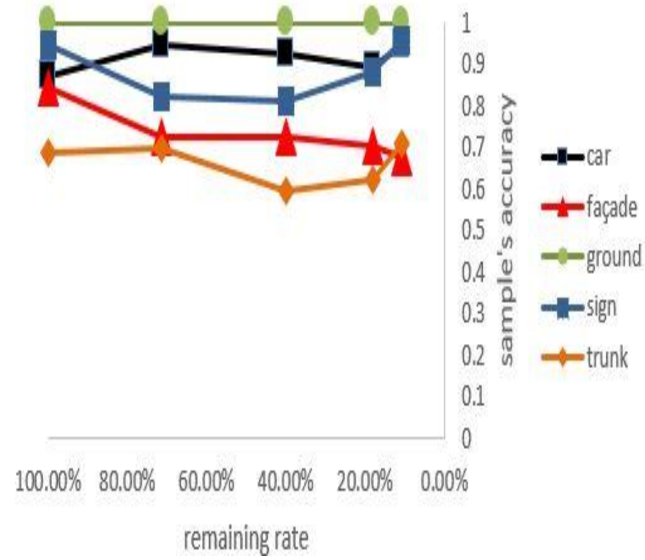
Table 4-8(a) Values of core measures from original-reduced data (façade) (b) Values of core measures based on original-reduced-reduced data (façade)

For method 6:

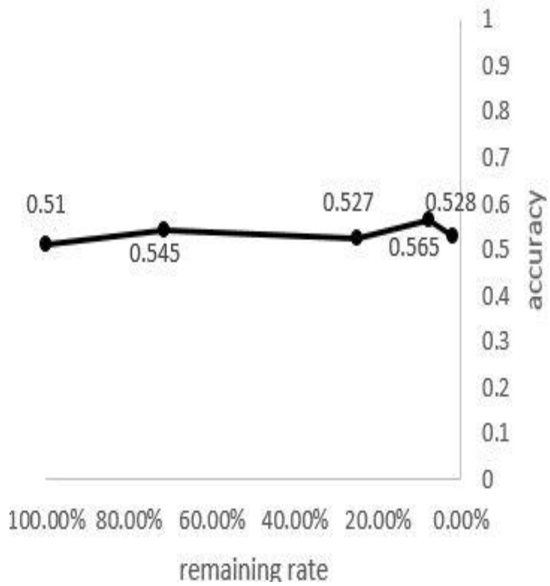
From Figure 4-19(a), there only exists five reduction levels to fine tune, because every time lowering one subdivision level, the large amount of points are deleted. From Table 4-9 and Figure 19(d), in the iterative point reduction, segmentation accuracy remain the same, and there is a few differences between reduced datasets among these values of core measures. From Figure 4-20, there is almost no difference of point distribution between these two datasets. Since method 6 is to remove points based on "Octree, so after direct point reduction once, the size of "Octree" has been determined, then regardless of how many times running this method to reduce points based on the same size of "Octree", no points are deleted. Overall, point distributions from iterative point reduction with the same subdivision level remain the same.



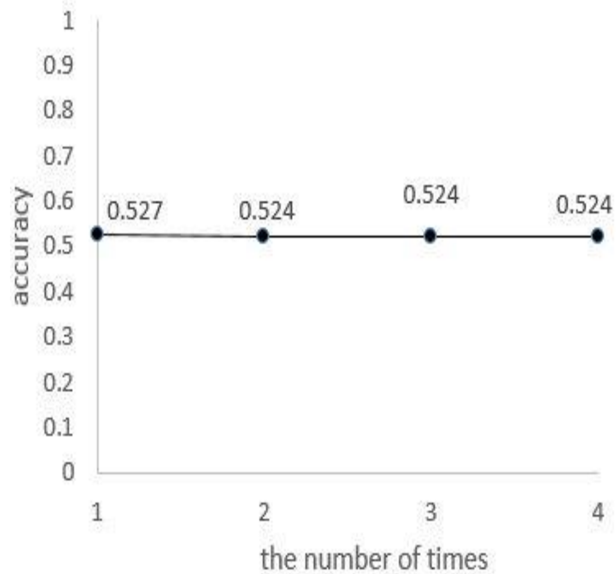
(a) Overall segmentation accuracy



(b) Every sample's segmentation accuracy



(c) Segmentation accuracy of direct point reduction dataset (one façade dataset)



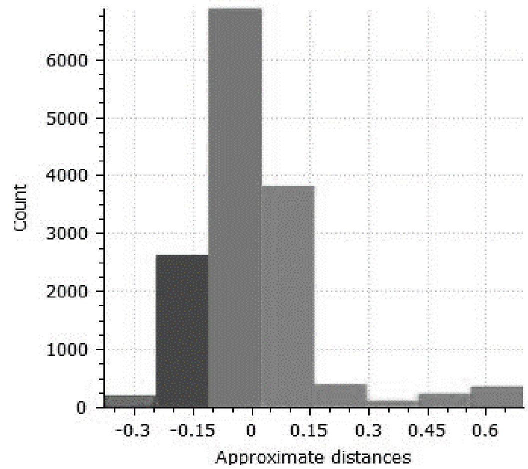
(d) Segmentation accuracy of iterative point reduction dataset (one façade dataset)

Figure 4-19 Segmentation accuracy

	Reduction Times	Reduction Parameter	Remaining rate	accuracy	OverSeg	UnderSeg	GroundSeg
(a)	1	8	25%	0.527	0.324	0.148	0.381
(b)	2	8	24.4%	0.524	0.361	0.115	0.471

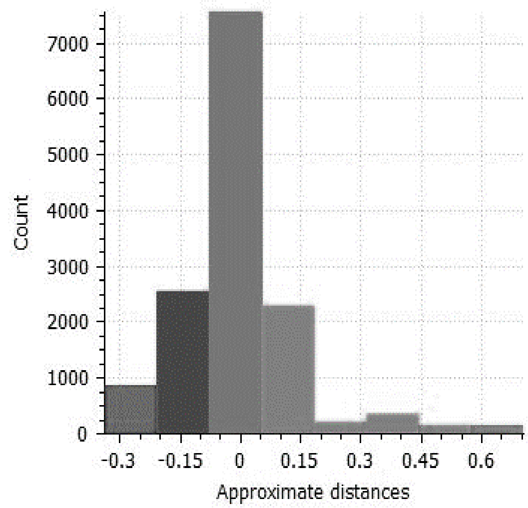
Table 4-9 (a) Values of core measures from original-reduced data (façade) (b) Values of core measures based on original-reduced-reduced data (façade)

(a) Segments within wall based on original-reduced dataset



(b) Distances from points to the fitted plane (original-reduced dataset)

(c) Segments within wall based on original-reduced-reduced dataset

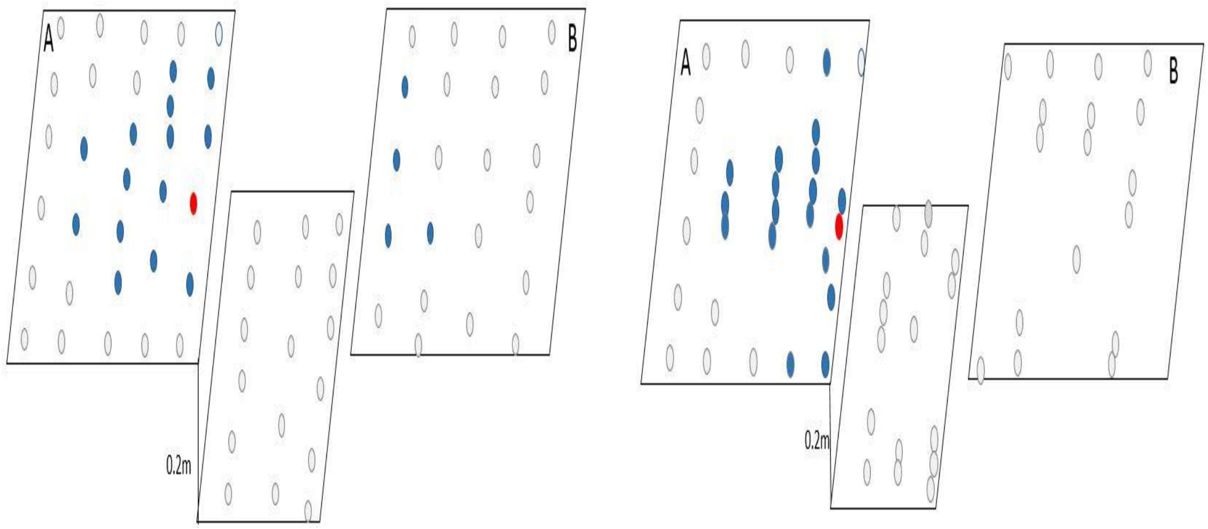


(d) Distances from points to the fitted plane (original-reduced-reduced dataset)

Figure 4-20 Visualization of point distribution on the wall

From results obtained from method 1 to method 6, segmentation accuracy of ground, trunk, sign, and car is stable and high at different point reduction levels. It seems that these objects can be analyzed for the influence of point reduction on surface growing. However, most of segments in the objects like car, sign and trunk are "non-planar", which are not suitable for assessing surface growing segmentation. Since structure of Façade is complex, which has more planar segments than that of other objects. So only select this façade data to do analysis about effect of point reduction on segmentation results.

In terms of segmentation accuracy of the façade, there exist two cases, one of cases is that segmentation accuracy of reduced data is better than original data. After point reduction, reduced data become more even than original data, which is shown in Figure 4-21 (a). So when seed points in flat parts want to grow, the seed point in plane A can find the suitable candidate points in plane B. Points in the flat part on the wall will be assigned into one segment, which lead to high accuracy results.



(a) (b)
Figure 4-21 Visualization of point distribution (example)

The other case is that segmentation accuracy become worse after point reduction. From Figure 4-21(b), even point density of reduced dataset is much lower than that of original dataset, resulting point distribution is still uneven. So when points in flat parts want to grow, the seed point in plane A can not find the suitable candidate points in plane B. So flat parts are still separated by the protruding portion.

For the time being, applicability of every method is evaluated on the basis of one façade dataset. Evaluation of each method is shown in Table 4-10. Two measures are used to determine every method's applicability: The range of segmentation accuracy and stability of every method.

Method	Direct point reduction (facade)		Iterative point reduction (facade)	
	Segmentation accuracy (range)	Stability	Segmentation accuracy (range)	Stability
1	[0.573, 0.744]	High	[0.573, 0.557]	Extremely high
2	[0.388, 0.845]	High	[0.388, 0.758]	High
3	[0.675, 0.797]	Low	[0.519, 0.675]	Low
4	[0.368, 0.654]	Low	[0.368, 0.7]	Low
5	0.76	Extremely High	0.394	Extremely High
6	[0.527, 0.565]	High	0.524	Extremely High

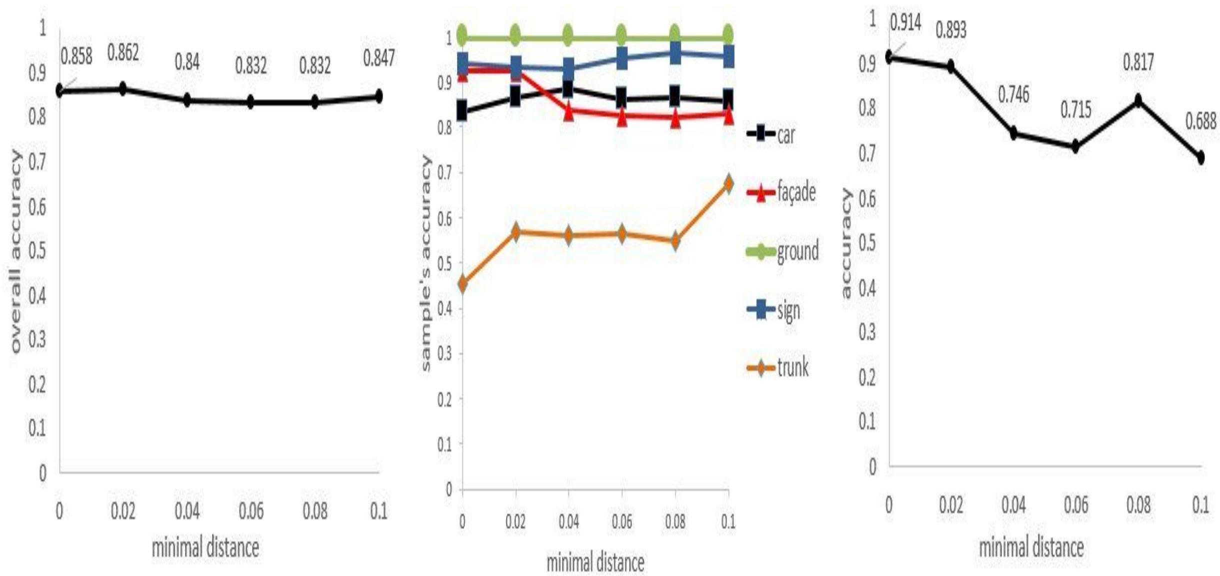
Table 4-10 Applicability of every method

For method 1, 2, 6, at the same point reduction level, segmentation results obtained by direct point reduction and iterative point reduction remain same. So when using these methods to reduce data, iterative point reduction has same behavior. Method 3 and 4 are much instable, even though use the same reduction parameter to reduce the same point cloud, segmentation results are different. Method 6 has limited values of reduction parameters to fine tune, which only lead to limited point reduction levels. In terms of segmentation accuracy of façade, users can get higher segmentation accuracy of façade data reduced by these six point reduction method than original data.

4.4 Analysis on effects of point reduction on segmentation of Enschede data

Like the part of analysis on effects of point reduction on segmentation of Paris data, parameters of surface growing segmentation stay unchanged, and "Number of neighbors in kd-tree" is set to 33. Only fine tune parameter of reduction parameter to get reduced datasets at different reduction levels so as to analyze what impact of point reduction has on Enschede dataset. After that, overall accuracy, every sample's accuracy and segmentation accuracy of one facade dataset are calculated.

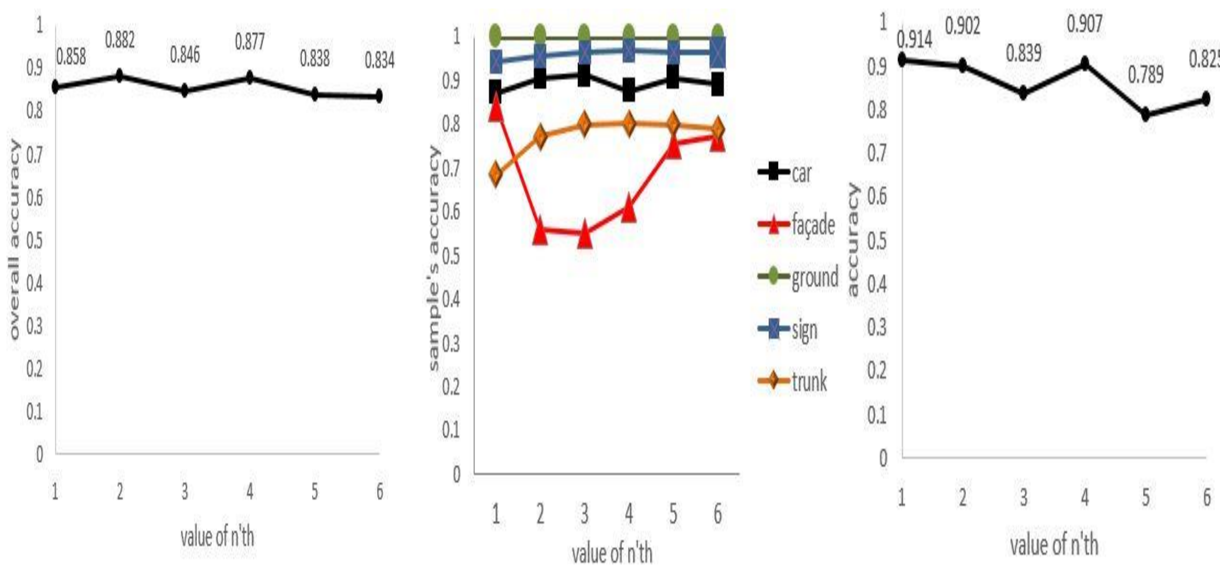
For method 1:



(a) Overall accuracy (b) Different sample's accuracy (c) Segmentation accuracy of one façade dataset

Figure 4-22 Segmentation accuracy

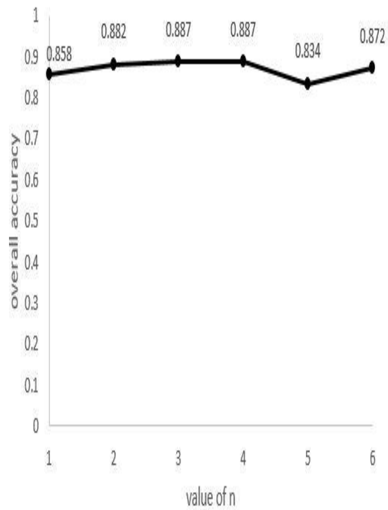
For method 2:



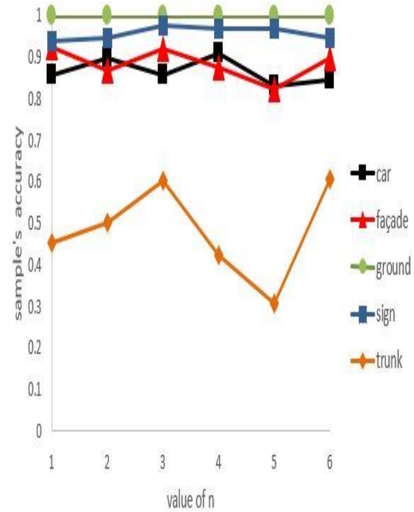
(a) Overall accuracy (b) Different sample's accuracy (c) Segmentation accuracy of one façade dataset

Figure 4-23 Segmentation accuracy

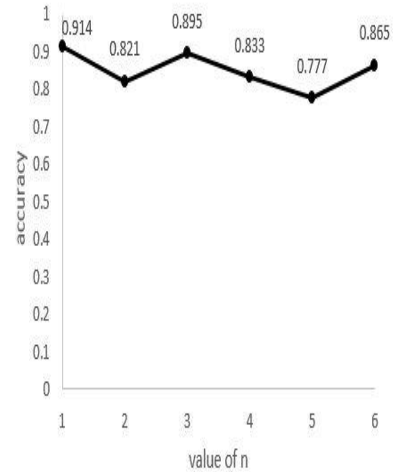
For method 3:



(a) Overall accuracy



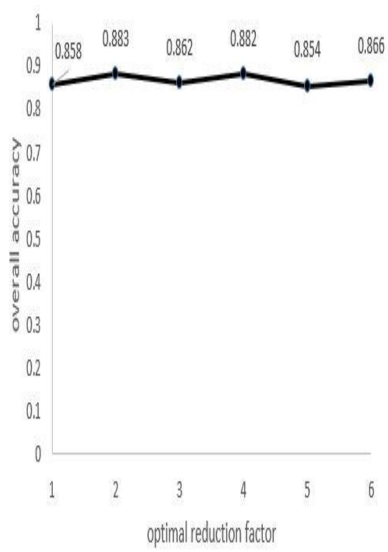
(b) Different sample's accuracy



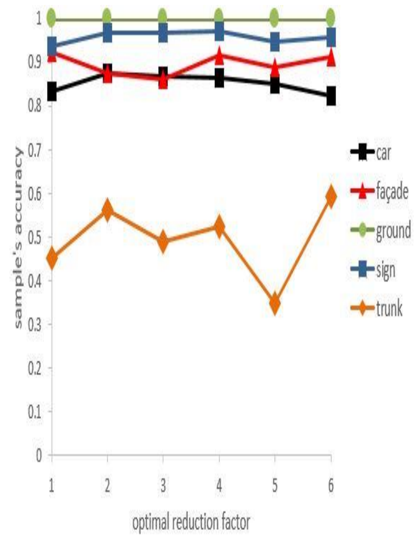
(c) Segmentation accuracy of one façade dataset

Figure 4-24 Segmentation accuracy

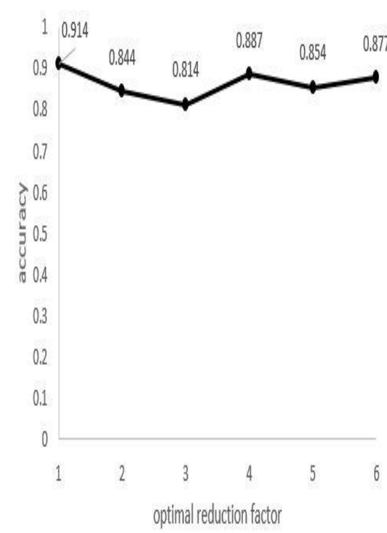
For method 4:



(a) Overall accuracy



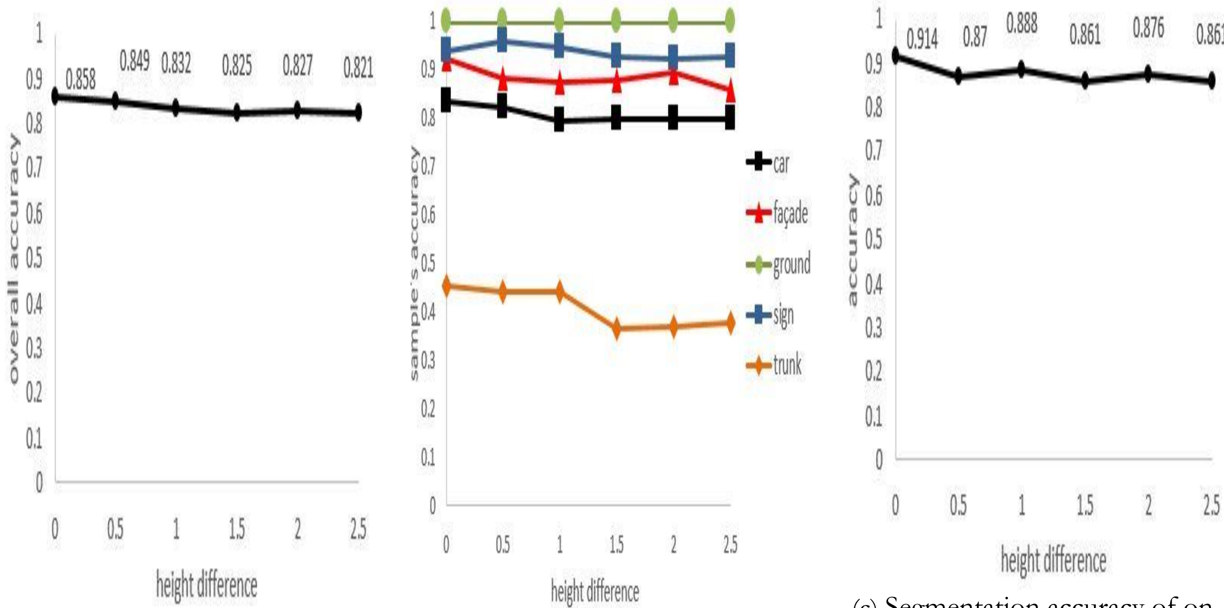
(b) Different sample's accuracy



(c) Segmentation accuracy of one façade dataset

Figure 4-25 Segmentation accuracy

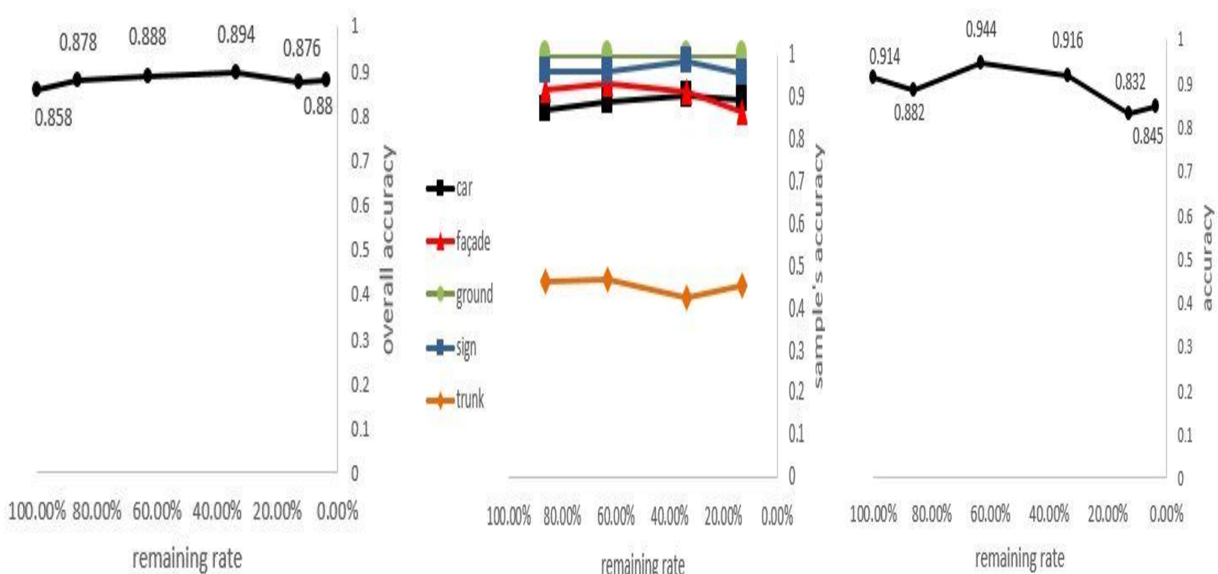
For method 5:



(a) Overall accuracy (b) Different sample's accuracy (c) Segmentation accuracy of on one façade dataset

Figure 4-26 Segmentation accuracy

For method 6:



(a) Overall accuracy (b) Different sample's accuracy (c) Segmentation accuracy of one façade dataset

Figure 4-27 Segmentation accuracy

From the segmentation results acquired by from method 1 to 6. After point reduction, overall segmentation accuracy of Enschede data is stable with high accuracy, also every sample's accuracy is high except segments in the trunk. Shape of trunk is irregular, which results in much over-segmentation. Therefor it is not suitable for assessing surface growing segmentation. Also the segments of car and sign is "non-planar", so they can not be used for assessment of surface growing segmentation. Segmentation accuracy of one façade is still high at different point reduction levels. So these point reduction methods do not change its point distribution a lot. In a word, all point reduction methods used in this study has good applicability applied for Enschede data.

5 Discussion

This chapter presents some limitations and problems which have not been tackled during the research.

5.1 Ground truth selection

Only surface growing segmentation method is applied in this research, which is well-used to detect planar segments. Quality of ground truths selection is determined by users, in this phase, the selected ground truths vary from person to person, with poor stability, especially for segments in car, trunk, and sign. The reason is these objects are complicated like cars, which completely consist of non-planar segments. So some selected ground truth segments in trunk, signs and cars cannot be regarded as absolute "planar segments", which is not suitable for assessment of the surface growing segmentation method. Therefor analysis of effects of point reduction/point distribution is mainly based on segments in façade. While in Paris data, structure of façade is more complicated than that in Enschede data because of prodruding elements on the wall. This is also a main reason why segmentation accuracy of façade in Enschede data is much higher than that of Paris data. As for the ground elements, they are completely planar and their structure is simple. Even though ground data is reduced by these point reduction methods, segmentation accuracy of reduced ground data is still 100%.

5.2 Surface growing parameters

Parameters of surface growing segmentation are empirical values, especially for Maximum distance to the surface. It has a big impact on segmentation results, because in the growing phase, it can determine whether to accept the candidate around seed points. In order to select approximate parameter of "Maximum distance to the surface", other parameter can be left default, only change this value, and make a trade-off between resulting segments in the Scenario containing various objects. Then selected the most suitable one by visual interpretation.

5.3 Quality assessment model

In this research, the quality assessment model presented in this research can evaluate segmentation accuracy of various objects, which is only used for scenario with ground truths. And values of core measures can provide details of results for users, which can help user to analyze resulting segments. In a word, the implementation of this model can meet users' requirement. However, it is not well-used for the large number of data, because ground truths are selected by users. If user want to get highly accurate evaluation of the large number of data, users should select ground truths as more as possible.

5.4 Evaluate applicability of every point reduction

In this study, six existing point reduction methods are used to reduce data. Two measures are defined to evaluate applicability of every point reduction method: Segmentation accuracy of facade and their Stability. Details of analysis is in section 4.3 and section 4.4. Method 5 based on minimum height difference has bad behavior when dealing with Paris data. The method 3, 4 is too randomized, users can't get stable segmentation results. Users can only get limit point reduction levels form method 6. Method 1 and 2 are stable, segmentation accuracy based on some reduced datasets obtained by these two methods is higher than that of original data, but the range of segmentation accuracy of method 2 is larger than that of method 1, so users may get very bad accuracy results from method 2. However, all these point reduction methods are well-used for Enschede data.

6 Conclusions and recommendations

This research aim at analyzing effects of point reduction on segmentation results of MLS data. Two datasets acquired by different types of scanner configurations and six existing point reduction methods are used in this research. Quality assessment model is proposed to evaluate segmentation accuracy in this research. Finally, distribution of distance from every point to the fitted plane and visualization of resulting segments are combined together to analyze what impact of point distribution\point reduction on segmentation results. This chapter is composed of two parts: the answer for every research question and some recommendations for further research.

6.1 Conclusion

Five research questions are proposed in this paper.

First and second questions were "What is the difference between point clouds acquired by laser scanning with different configurations?", "What is the effect of point distribution on the segmentation of MLS data?"

To answer these questions, Enschede and Paris data acquired by mobile laser scanning with different scanner configurations are used, which have different point distributions. For Paris dataset, in terms of segmentation accuracy of façade, it is so low due to its uneven point distribution. For Enschede data, in terms of segmentation accuracy of façade, it is very high because of its even point distribution. So we can conclude that even point distribution can lead to high segmentation accuracy. On the contrary, uneven point distribution can lead to low segmentation accuracy.

The third, fourth, and fifth questions are "What is the influence of point reduction to the segmentation results in terms of quality? /How can the quality of the resulting segmentation be assessed (with respect to the point distribution/reduction)/ Evaluate applicability of every point reduction method and to see whether to propose a point reduction method to get a higher quality segmentation results?"

To answer these three questions, six existing point reduction methods in PCM and CloudCompare "subsample" tool are used in this study. By quality assessment model, AverAccuracy, Overseg, UnderSeg, and GroundSeg can be calculated, segmentation accuracy of all reduced datasets can be assessed.

In terms of analysis about what impact of point reduction has on segmentation results of Paris data, only select one façade data to do further analysis, there exist two cases: For Paris data, one of cases is that after point reduction, point distribution become more even than original data, so higher segmentation accuracy can be obtained. The other one is that after point reduction, point distribution is still uneven, segmentation accuracy is still bad. In the evaluation of applicability of every reduction method on Paris data, from Table 4-10, considering segmentation accuracy of segments in façade and its stability, space-based point reduction (Method 1) is stable and high. So space-based point reduction is selected the most suitable one. At least the same level, we can get even point distribution from space-based point reduction, so there is no need to propose a new method to reduce Paris data.

For Enschede data, in terms of segmentation accuracy of the façade dataset, it is very high because of its even point distribution. And after point reduction, segmentation accuracy of the façade is still high, so these point reduction methods have good performance. There is no need to propose a new method to reduce Enschede data.

So we can conclude that: For even point distribution, point cloud reduced by these point reduction methods, segmentation accuracy is still high. For uneven point distribution, if suitable point reduction is selected, resulting point distribution will be even, which can lead to high segmentation accuracy.

6.2 Recommendation

Evaluating point clouds' segmentation accuracy is still a long-term research. In this study, the quality assessment model only works with scenario with ground truths, and ground truth segments have to be selected by users.

It is recommend that users should select segments in the façade and ground to do assessment of surface growing segmentation. On the contrary, some objects like cars, signs, trunks, their structure is complicated, segments in these objects are not planar, which is not suitable for assessment of surface growing segmentation. Here six point reduction methods have been analyzed, so user can choose the most suitable one according to data. A variety of point clouds are put into use in recent years, so researchers still need to propose more useful point reduction methods according to applications.

List of references

Clinton, N., Holt, A., Scarborough, J., Yan, L., & Gong, P. (2010). Accuracy Assessment Measures for Object-based Image Segmentation Goodness. *Photogrammetric Engineering & Remote Sensing*, *76*(3), 289–299. http://doi.org/10.14358/PERS.76.3.289

Gonzalez-Jorge, H., Solla, M., Armesto, J., & Arias, P. (2012). Novel method to determine laser scanner accuracy for applications in civil engineering. *Optica Applicata*, *42*(1), 43–53. http://doi.org/10.5277/oa120104

Khoshelham, K., Altundag, D., Ngan-Tillard, D., & Menenti, M. (2011). Influence of range measurement noise on roughness characterization of rock surfaces using terrestrial laser scanning. *International Journal of Rock Mechanics and Mining Sciences*, *48*(8), 1215–1223. http://doi.org/10.1016/j.ijrmms.2011.09.007

Lee, N. S. (2009). Data Reduction of point clouds acquired by airborne laser scanning Data Reduction of point clouds acquired by airborne laser scanning.

Moenning, C., & Dodgson, N. (2003). A new point cloud simplification algorithm. In *Proc. Int. Conf. on Visualization, Imaging Processing* (pp. 8–10). Benalmádena.

Möller, M., Lymburner, L., & Volk, M. (2007). The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation*, *9*(3), 311–321. http://doi.org/10.1016/j.jag.2006.10.002

Ni, N., Chen, N., & Chen, J. (2014). Accuracy evaluation of segmentation for high resolution imagery and 3D laser point cloud data. In A. G. Tescher (Ed.), *SPIE Optical Engineering + Applications* (p. 921728). International Society for Optics and Photonics. http://doi.org/10.1117/12.2060306

Pu, S., Rutzinger, M., Vosselman, G., & Oude Elberink, S. (2011). Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), S28–S39. http://doi.org/10.1016/j.isprsjprs.2011.08.006

Puente, I., González-Jorge, H., Martínez-Sánchez, J., & Arias, P. (2013). Review of mobile mapping and surveying technologies. *Measurement*, *46*(7), 2127–2145. http://doi.org/10.1016/j.measurement.2013.03.006

Rutzinger, M., Pratihast, A., Elberink, O., & Vosselman, G. (2010). Detection and Modelling of 3D Trees From Mobile Laser Scanning Data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(5), 520–525. Retrieved from http://www.mendeley.com/catalog/detection-modelling-3d-trees-mobile-laser-scanning-data/

Vosselman, G. (2013). Point cloud segmentation for urban scene classification. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-7/W2*(November), 257–262. http://doi.org/10.5194/isprsarchives-XL-7-W2-257-2013

Vosselman, G., Gorte, B. G. H., Sithole, G., & Rabbani, T. (2004). Recognising strucutre in laser scanner point clouds. *Remote Sensing and Spatial Information Sciences*, *32*(5), 33–38. http://doi.org/10.1002/bip.360320508

Weidner, U. (2008). Contribution to the assessment of segmentation quality for remote sensing applications. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *37*(B7), 479–484. Retrieved from http://www.mendeley.com/catalog/contribution-assessment-segmentation-quality-remote-sensing-applications/

Weinmann, M., Jutzi, B., Hinz, S., & Mallet, C. (2015). Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 286–304. http://doi.org/10.1016/j.isprsjprs.2015.01.016