Data Reduction of point clouds acquired by airborne laser scanning

Nam Sik Lee
March, 2009
Data Reduction of point clouds acquired by airborne laser scanning

by

Nam Sik Lee

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialization: Geoinformatics.

Thesis Assessment Board

Chairman: Prof. Dr. A. Stein
External examiner: Dr. R.C. Lindenbergh
Supervisor: Dr. M. Gerke
Second supervisor: Prof. Dr. M.G. Vosselman (RTL)
Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.
Abstract

The modern development of laser scanning devices makes it possible to achieve detailed descriptions of the surveyed surfaces with a huge amount of points. It is clear that we can more accurately describe their properties with more points. However, all the acquired points are not necessary for the proper processing and for the best usage by all applications. Hence, we need methods for data reduction which can remove redundant points from point clouds without losing important information.

This research tries to develop a new point-based simplification method being able to remove the points which contribute less to the surfaces. In the proposed method, every single point will be evaluated whether to be remained or removed based on the demands on representing the surfaces, which will be called significance value in this paper. The criteria will be embodied as two parameters: Sum of height difference and change of normal vector with respect to a locally fitted plane. Significance values are calculated by two parameters and points to be remained are selected by applying threshold. In the research, two ways of finding threshold values by each parameter will be proposed: one is to specify the ranges on the visualized values, and the other is to use tables classified on the basis of specific surface properties.

The proposed method was tested with two data sets: building and terrain acquired by airborne laser scanning. The reduced data in the terrain shows better result by the proposed accuracy assessment method, which is the comparison of DEM subtraction, rather than in the building. That is because the building is composed of many points with big height jumps which can cause big errors to the rasterized DEM. So, this result should be interpreted only as the adopted accuracy assessment method may not be appropriate for the building, instead quite useful for terrain when making DTM.

This research proposes appropriate parameters which can reflect the properties of points acquired by airborne laser scanning, and quantitative accuracy assessment method suitable for terrain data. In addition, the proposed selection method is quite flexible for users to select the points interactively by adjusting one of two threshold values on their interests.

Keywords: Airborne laser scanning, Point cloud, significance value, neighbourhood, height difference, normal vector, DEM.
Acknowledgements

It was my wonderful experience in my life studying here in ITC (International Institute for Geo-
information Science and Earth Observation) in Enschede, the Netherlands. I could open my eyes to the
worlds as well as expand my knowledge on new fields in Geoinformatics. This was only possible with
the help of my country and many people.

I would like to express my first thanks to the Republic of Korea Army. I could have this experience
with their full supports for studying and living here. I am deeply indebted to ROKA for giving this
opportunity. I am always proud of me being a military officer in ROKA, and I will do my best to
contribute to them after coming back.

I want to extend my thanks to my two supervisors. Dr. Markus Gerke, my first supervisor, gave me his
unlimited and continuous instructions during my whole research periods. No other supervisors could
do better than he did to support the student. He was always available even in the weekends and eager
to guide me with his detailed instructions. I could learn from him what a research is and how scientific
paper should be written. Prof. Dr. M.G. Vosselman, my second supervisor, was really professional and
encouraged me to go in right directions. I could keep continuing my research with his constructive and
clear advices. I was very lucky during my research periods since I had these really professional and
cool two supervisors

I also appreciate for my ass GFM classmates. Although we have different cultures, languages, and
different colours, we could understand each other and could be good friends. We were always together
when studying and playing. Especially, I would like to give my special thanks to three classmates:
Wanggyu Jeon (Korean), Zhou Liang (Chinese), Thapa Anisha (Nepalese). We were always
occupying the first row in the classroom, and helping and encouraging each other. I could keep
continuing my research with the help of Jeon’s professional knowledge, Zhou’s smart brain, and
Anisha’s fluent English. Thank them for being my classmates in ITC.

Last but not least, I would like to give my best thanks to my wife. I could not have finished my thesis
successfully without her devotion to support me and my family. She always put the first priority on her
family members before herself, especially to take care of my two daughters. I could concentrate on my
studying only with her love and full supports. I am really happy to have this wonderful and beautiful
wife. To my love, Soyoung, All this work is dedicated to you. In addition, I also want to express my
unlimited love to my two lovely daughters. They had a hard time at first when we came here due to the
language and cultural differences. However, they could enjoy their school life and living in the
Netherlands for the last half of the period. I appreciate for them to adjust themselves well to the new
environments without complaints.

Thanks and love all of you...
# Table of contents

1. Introduction ........................................................................................................................................... 1  
   1.1. Motivation and problem statement ........................................................................................................ 1  
   1.2. Research Identification ........................................................................................................................... 2  
   1.2.1. Research objectives ............................................................................................................................... 2  
   1.2.2. Research questions ............................................................................................................................... 2  
   1.2.3. Innovation aimed at ............................................................................................................................. 2  
   1.3. Thesis structures ..................................................................................................................................... 3  
2. Review of data reduction methods of point clouds .............................................................................. 4  
   2.1. Introduction ........................................................................................................................................... 4  
   2.2. Clustering .............................................................................................................................................. 4  
   2.2.1. Model simplification through refinement (Brodsky and Watson, 2000) ........................................... 4  
   2.2.2. Efficient adaptive simplification of massive meshes (Shaffer and Garland, 2001) ....................... 6  
   2.2.3. Multi-resolution 3D approximations for rendering complex scenes (Rossignac and Borrel, 1993) ................................................................................................................................. 8  
   2.3. Particle simulation ................................................................................................................................. 9  
   2.3.1. Re-Tiling polygonal surfaces (Turk, 1992) ............................................................................................ 9  
   2.4. Iterative simplification ............................................................................................................................ 10  
   2.4.1. Point set surfaces (Alexa et al., 2001) ............................................................................................... 10  
   2.4.2. Progressive point set surfaces (Fleishman et al., 2003) ................................................................... 12  
   2.4.3. Surface simplification using quadric error metrics (Garland and Heckbert, 1997) ....................... 13  
   2.4.4. Point cloud representation (Linsen, 2001) ......................................................................................... 14  
   2.4.5. Meshfree thinning of 3D point cloud (Dyn and Iske, 2007) .............................................................. 16  
   2.5. Summary and requirements for a new method ..................................................................................... 18  
   2.5.1. Analysis on the introduced literatures ................................................................................................. 18  
   2.5.2. Requirements for a new method ........................................................................................................ 20  
3. Proposed method for data reduction .................................................................................................... 21  
   3.1. Pre-processing ..................................................................................................................................... 22  
   3.1.1. Defining the neighbourhood ............................................................................................................... 22  
   3.1.2. Estimation of normal vector ............................................................................................................... 23  
   3.2. Derivation of parameters for calculating significane value ................................................................. 24  
   3.2.1. Calculating sum of height difference within neighboured points .................................................. 24  
   3.2.2. Calculating change of normal vectors within neighboured points .............................................. 27  
   3.3. Data reduction ..................................................................................................................................... 28  
   3.4. How to find threshold values ................................................................................................................. 29  
   3.4.1. Visualization of calculated values ..................................................................................................... 29  
   3.4.2. Differences of DEM by original and reduced data ............................................................................ 30  
      3.4.2.1. RMSE as a measure of accuracy assessment .............................................................................. 30  
      3.4.2.2. Subtracting one DEM from the other DEM .............................................................................. 31  
   3.4.3. Visual check ...................................................................................................................................... 32  
   3.5. Accuracy assessment ............................................................................................................................ 32  
   3.6. Conclusion of the proposed method ..................................................................................................... 32  
   3.6.1. Summary .......................................................................................................................................... 32  
   3.6.2. Comparison with other methods ........................................................................................................ 33  
4. Implementation and results .................................................................................................................. 34
4.1. Study area and input data ................................................................. 34
4.2. Calculating significance value .......................................................... 34
  4.2.1. Test with a building ........................................................................ 34
    4.2.1.1. Significance value by sum of height difference ......................... 34
    4.2.1.2. Significance value by change of normal .................................... 36
    4.2.1.3. Analysis on significance values of the building ......................... 37
  4.2.2. Test with terrain ............................................................................ 37
    4.2.2.1. Significance value by sum of height difference ......................... 38
    4.2.2.2. Significance value by change of normal .................................... 39
    4.2.2.3. Analysis on the significance values of the terrain ....................... 40
4.3. The result of data reduction by two parameters ...................................... 40
  4.3.1. Data reduction .............................................................................. 40
  4.3.2. Analysis on the result of data reduction and RMSE ......................... 40
  4.3.3. Visual check .................................................................................... 43
5. Discussion ............................................................................................... 46
  5.1. On neighbourhood problem ............................................................... 46
  5.2. Difference of DEM as a measure of accuracy assessment for building? .... 47
  5.3. On the threshold values ..................................................................... 48
6. Conclusions and recommendataions .......................................................... 49
  6.1. Conclusions ....................................................................................... 49
  6.2. Recommendations ............................................................................. 50
Reference .................................................................................................... 51
Appendix A: Remained rate and RMSE ......................................................... 53
Appendix B: Visualization of remaining points ............................................. 57
List of figures

Figure 2-1: local neighbourhood (a) and covariance analysis (b), (Pauly et al., 2002) .......................... 6
Figure 2-2: Hierarchical clustering based on the BSP-Tree structure, where thickness of the lines indicates the level of BSP tree, (Pauly et al., 2002) ........................................................................... 7
Figure 2-3: Overview of the simplification process, (Rossignac and Borrel, 1993) ............................... 8
Figure 2-4: Removing a vertex(R) from a mutual tessellation, (Turk, 1992) ......................................... 10
Figure 2-5: The MLS projection procedure, (Alexa et al., 2001) .......................................................... 11
Figure 2-6: Refinement process, (Fleishman et al., 2003) .................................................................... 12
Figure 2-7: Edge contraction, (Garland and Heckbert, 1997) ............................................................... 13
Figure 2-8: Neighbourhoods of a point $p$, (Linsen, 2001) ............................................................... 14
Figure 2-9: Illustration of surface features, (Linsen, 2001) ................................................................. 15
Figure 2-10: Result of data reduction (42% of original), (Linsen, 2001) ................................................ 16
Figure 2-11: Illustration of procedures ............................................................................................... 17
Figure 2-12: Result of data reduction, (Dyn and Iske, 2007) ............................................................... 17
Figure 3-1: Workflow applied in the research ....................................................................................... 21
Figure 3-2: Neighbourhood in 2 dimensions and 3 dimensions ............................................................ 22
Figure 3-3: A normal to a surface or tangent plane at point $P$ ............................................................ 23
Figure 3-4: Sum of height differences .................................................................................................. 25
Figure 3-5: Necessity for incorporating maximum height differences .................................................. 26
Figure 3-6: Maximum height difference .............................................................................................. 26
Figure 3-7: Change of normal vectors ................................................................................................ 27
Figure 3-8: Angle between two unit normal vectors ............................................................................ 28
Figure 3-9: Methodology of data reduction ......................................................................................... 28
Figure 3-10: Histogram equalization .................................................................................................. 30
Figure 3-11: Procedure of DEM generation ....................................................................................... 31
Figure 3-12: Selection of cell size ....................................................................................................... 31
Figure 3-13: Procedure for Subtraction of two DEMs ....................................................................... 32
Figure 4-1: Town hall of Enschede ..................................................................................................... 34
Figure 4-2: Distribution of sum of height difference ........................................................................... 35
Figure 4-3: Visualization of sum of height difference ........................................................................ 35
Figure 4-4: Distribution of change of normal .................................................................................... 36
Figure 4-5: Visualization of change of normal ................................................................................... 36
Figure 4-6: University Twente in Enschede ....................................................................................... 37
Figure 4-7: Distribution of sum of height difference ........................................................................... 38
Figure 4-8: Visualization of sum of height difference ........................................................................ 38
Figure 4-9: Distribution of change of normal .................................................................................... 39
Figure 4-10: Visualization of change of normal .................................................................................. 39
Figure 4-11: Results by applying different height differences (example) ........................................... 41
Figure 4-12: Results by applying different change of normal (example) .......................................... 42
Figure 4-13: Surface representations ................................................................................................ 43
Figure 4-14: Visualization of classification table ............................................................................... 45
Figure 5-1: Comparison of neighbourhood ........................................................................................ 46
Figure 5-3: Visualization of differences of DEM subtractions (example) ........................................... 47
List of tables

Table 2-1: Summary of the algorithms for data reduction ................................................................. 19
Table 3-1: The range of initially calculated values by two parameters ........................................... 29
Table 3-2: The table of classification ................................................................................................. 29
Table 4-1: Data reduction and accuracy in building (from 37,875 points) ........................................ 41
Table 4-2: Data reduction and accuracy in terrain (from 45,126 points) .......................................... 41
Table 4-3: Selection of two threshold values with the smallest \textit{RMSE} ........................................... 42
Table 4-4: Classification tables for Building ...................................................................................... 44
Table 4-5: Classification table for terrain ......................................................................................... 44
1. Introduction

1.1. Motivation and problem statement

Laser scanning technology has been adopted as a main method for acquiring geometric data of physical surfaces in recent years. Accordingly, laser scanning devices has been developed rapidly; making it possible to produce enormous amount of points on the objects with high precision (Moenning and Dodgson, 2006). For example, commercial airborne laser scanners boarded on a fixed wings aircraft can operate on up to 5000 m height; measuring the points up to frequencies 250,000 and even higher with a few cm accuracies. This leads to a typical point density of several or even higher points per m². (Kobler et al., 2007)

The increasing point density enables achieving a detailed description of the surveyed surfaces and provides wealth of information on manmade objects or terrains (Filin, 2004). However, these initial data generally consist of considerable amount of redundant points which should be removed for efficient usage by other applications. For example, in the case of 3D building modelling as a part of 3D city model, the points to be used for modelling will be several corner points, while most points within the same plane are redundancies and thus to be removed, even though the high point density is still required in some applications like visualization or interpretation of surface roughness. In addition, processing the huge amount of data requires huge computation and memory demands.

A variety of methods have been developed for reducing the amount of initial points, commonly being referred to as point clouds simplification, decimation, or thinning. With the increasing availability of point-based modelling, multi-resolution, and visualization techniques, the point-based simplification has become a more attractive approach than the polygonal mesh-based simplification (Moenning and Dodgson, 2003). An extensive survey of mesh-based simplification methods can be found in (Heckbert and Garland, 1997, Gotsman et al., 2002). In addition, point based simplification is more efficient and less memory demanding since no mesh data structures like topological and geometric features need to be preserved (Dey et al., 2001).

However, there is no global optimum suitable for every application in the existing point based simplification methods. For example, the reduction methods developed in the real-time applications put more emphasis on the efficiency rather than accuracy, while the methods developed in visualization applications try to describe the surfaces more accurately rather than efficiently. Each method has its specific pros and cons. Hence, the choice of method always depends on the intended application.

In addition, most of the existing simplification methods have been developed in the computer graphic fields with terrestrial laser data which is composed of continuous points on the objects. However, airborne laser scanning data has different properties compared to terrestrial laser scanning data.
Basically, the objects of airborne laser scanner are located in the large area and taken from the top viewing, although some points can be taken from oblique viewing. Hence, the data are composed of many points with big height jumps especially in the urban area. However, there is no method developed to handle these specific properties of airborne laser scanning data. This research is aimed at developing appropriate method for these airborne laser scanning data.

1.2. Research Identification

1.2.1. Research objectives

The overall objective of research is to develop a new method which can remove non-significant points from dense point clouds acquired by airborne laser scanning according to the calculated significance values. The way of calculating significance value of point should be defined in this research. This can be described as follows:

- To find a way of determining the significance value of every point based on the pre-defined parameters.
- To find out the impacts of different properties of data on the results.

1.2.2. Research questions

To achieve above objectives, the following research questions should be answered:

- How can parameters be defined in terms of surfaces representation?
- How can the significance value of each point be determined based on the pre-defined parameters?
- What is the impact of different types of surfaces on the results?
- What is the impact of different point density on the results?
- Can a method be found that removes all points except for building corner points?

1.2.3. Innovation aimed at

The idea of using significance value, sometimes under other terminologies like entropy of the points (Linsen, 2001) or contributions of the points to the surfaces (Alexa et al., 2001), as a criterion to remove points is not new (Dyn and Iske, 2007). However, these existing methods calculate the value of each point by their own parameters and remove the least valued point iteratively. In other words, the calculations are done in each step iteratively when removing the least valued point, accordingly resulting in different values in each step. However, the proposed method calculates the value of each point at one scan and applies the reduction method directly to the calculated values of all points. In addition, the proposed method is designed to handle the airborne laser scanning data, while the other existing methods have been developed for terrestrial laser scanning data. So, innovations aimed at in this research are:

- Definition of the most appropriate parameters for representing surfaces by airborne laser scanning data.
- Development of formula calculating significance values of all points by pre-defined parameters.
- Proposal of meaningful quantitative accuracy assessment method on the reduced point set.

1.3. Thesis structures

This research is divided into six main chapters. Chapter 1 introduces the background of this research including research objectives and questions. A chapter 2 reviews different method of reducing data found in literatures and derives the necessity of this research. In chapter 3, two parameters will be derived for computing significance value; the methods of reducing points and accuracy assessment will be proposed. Chapter 4 shows the results by applying the proposed method and compares the results in terms of accuracy. In chapter 5, several issues faced during research are discussed together with some suggestions. Chapter 6 summaries this research with brief answers to each research question and presents some recommendations for further developments.
2. Review of data reduction methods of point clouds

2.1. Introduction

There exist several different methods for point-based simplification, which are categorized into three classes and will be explained one by one in this chapter (Dyn and Iske, 2007, Pauly et al., 2002). The three main categories in point based simplifications are clustering, particle simulation, and iterative simplification method.

The first category is clustering method which splits the point clouds into a number of sub-sets and replaces them with representative points. The key concern in this method is how to split the 3D point clouds into sub-sets. Two general approaches for clustering method were proposed by (Pauly et al., 2002), which are a hierarchical surface clustering and an incremental surface clustering method. Two hierarchical clustering algorithms and one incremental clustering algorithm will be introduced here.

The second category is particle simulation method which computes the sampling positions with the desired number of points by using the inter-particle repelling forces. Normally, point positions are restricted to the local surfaces defined by individual polygons. There have not been many researches on this method. So, one algorithm published in 1992 will be introduced here.

The third category is the iterative simplification method which reduces the number of points based on certain criteria iteratively. Basically, most algorithms in this method try to define local planes or local surfaces to calculate value of certain points. To be more precise, the main concern of these methods is how to define the relevant reference domain and how to formulate the criteria to calculate values of points. After calculating values, the algorithms try to find and remove the least important points iteratively. Several different approaches of defining the criteria can be recognized in section 2.4.

For each approach presented here, the general description of the methodology will be highlighted and an evaluation will also be included if possible. In section 2.5, the general comparison of three methods will be given, and the requirements of a new method will be explained.

2.2. Clustering

2.2.1. Model simplification through refinement (Brodsky and Watson, 2000)

This algorithm, which is called $R$-simp, consists of three stages to reduce the amount of points: initialization, simplification, and post processing. In initialization stage, global face ($gfl$) and vertex ($gvl$) lists, and eight initial clusters are created by partitioning the model using three axis-aligned planes which are positioned in the middle of model’s bounding box.
In simplification stage, simplification process can be preceded on the basis of calculated face normal variations within the initial eight clusters. The simplification stage can be sub-divided into several steps as follows;

1. Choose the cluster which has the most face normal variation

Normal variation ($nv$) is defined to be sensitive to the size, so the complement of $cp$ is used so that $nv$ increases as the face normal variation increases;

$$nv = \sum_i \frac{a_i}{M} (1 - cp)$$  \hspace{1cm} (2.1)

Where,

$N$ = number of faces in the cluster

$M$ = number of faces in the model

$a_i$ = area of face $i$

$cp$ = face normal variation measure

($cp = \frac{\sum a_i}{\sum a_i}$, area-weighted mean of all face normal)

In addition, normal variation in the chosen cluster is described by covariance matrix in PCA (Principle Component Analysis) (Jolliffe, 2002, Pauly et al., 2002). For example, local surface of sample point $p$ can be estimated by eigenanalysis of covariance matrix within a local neighbourhood. The 3x3 covariance matrix $C$ for a sample point $p$ is given by

$$C = \begin{bmatrix} \vec{p}_1 - \vec{p} \\ \vdots \\ \vec{p}_k - \vec{p} \end{bmatrix}^T \begin{bmatrix} \vec{p}_1 - \vec{p} \\ \vdots \\ \vec{p}_k - \vec{p} \end{bmatrix}, \vec{p}_i \in N_p, \hspace{1cm} (2.2)$$

Where, $\vec{p}$ is the centroid of the neighbours $\vec{p}_i$ of $p$

The eigenvector problem can be explained by $C \cdot V_\lambda = \lambda_i \cdot V_\lambda$, $\lambda_i \in \{0,1,2\}$. Since $C$ is a 3x3 matrix with symmetric and positive semi-definite components, all eigenvalues $\lambda_i$ will be real-valued and corresponding eigenvectors $V_i$ form an orthogonal frame. The $\lambda_i$ measures the variation of the $p$, along the directions of corresponding eigenvectors, and the total variation from its center of gravity are given by $\sum_{i \in np} |p_i - \vec{p}|^2 = \lambda_0 + \lambda_1 + \lambda_2$. Here, assuming that $\lambda_0 \leq \lambda_1 \leq \lambda_2$, the largest eigenvalue and corresponding eigenvector in the covariance matrix represent the mean normal of the surface, the second and third largest eigenvalue and corresponding eigenvector represent the directions of
maximum and minimum curvature respectively. So, $\lambda_0$ quantitatively describes the variation along the surface normal, while $\lambda_1$ and $\lambda_2$ span the tangent plant $T(x)$ as is seen in Figure 2-1 (b).

![Figure 2-1: local neighbourhood (a) and covariance analysis (b), (Pauly et al., 2002)](image)

- Split the cluster based on the amount and direction of the face normal variations. In this step, the number of sub-clusters to be created, which are two, four and eight, is determined according to the ratio of between three eigenvalues. The position and orientation of splitting planes are determined by using eigenvectors corresponding to each eigenvalue. Sub-clusters are created by partitioning the vertices in a cluster.

- Compute the amount of face normal variation in each of the sub-clusters.

- Iterate until the required number of clusters is reached.

In post processing stage, for each cluster left, the representative vertex (centered point) is computed, and the model is re-triangulated. The location of representative vertex for each cluster is computed to minimize the summed distance from the planes containing the cluster’s face. The re-triangulation is done by using $gfl$ (global face list) and $svl$ (simplified vertex list) in which any face referencing three different vertices in $svl$ (simplified vertex list) is retained and output to the $sfl$ (simplified face list). All other faces have degenerated into lines or points and are discarded.

This algorithm is said to be well suited for simplification of large models in interactive environments since it can create continuous level of detail hierarchies with the curvature (face normal variation) information which was used in the simplification process. The process will proceed in a way that cells containing little curvature are to be split less than cells containing more curvature. The author said that this algorithm’s time guarantees and quality/speed tradeoffs make it ideal for use in interactive application.

2.2.2. Efficient adaptive simplification of massive meshes (Shaffer and Garland, 2001)

This algorithm consists of three steps for data reduction. In the first step, the input mesh is quantized using a uniform grid. And then, vertices of the model are partitioned by using the accumulated geometric (surface) information in the form of quadrics and dual quadrics in the second step. In the final step, the original mesh is simplified using this partition information. The spatial partitioning
technique adopted in this algorithm is based on BSP-trees structure in which simplification proceeds from a coarse to fine level (Samet, 1990, Pauly et al., 2002).

In the quantizing step, quadrics are accumulated in the cells produced by a regular grid partition of the space. For each face in a mesh defining a plane which satisfies the equation $n v^T + d = 0$, where $n$ is unit normal vector, the squared distance of a vertex $v$ to this plane can be represented by using a quadric $Q$. So, the sum of squared distances to a set of planes as a form of quadric is obtained from a single quadric which is the sum of quadrics defined by each individual plane. In addition, curvature information associated with set of planes is also encoded in the quadric matrix by using eigenanalysis. Just as quadric metric, the distance from a plane to a set of points can be represented as dual quadric. Finally, each vertex in the mesh generates a dual quadric which is added to the containing grid cell, while each face will generate a quadric which is then added to the three cells.

![Hierarchical clustering based on the BSP-Tree structure](image)

Figure 2-2: Hierarchical clustering based on the BSP-Tree structure, where thickness of the lines indicates the level of BSP tree, (Pauly et al., 2002)

In spatial partitioning step, the leaf with the largest primal quadric error (left side of the split plane in Figure 2-2 is chosen to be split, and a representative point $v_2$ for each quantized cell is chosen to minimize the primal quadric error. These representative points are then inserted into a BSP-tree. The dual quadric associated with a leaf is used to determine a splitting plane. Once the splitting plane is determined, eigenvectors and eigenvalues of the point set using covariance matrix associated with that leaf are extracted to be used to position the splitting plane. The detailed procedure is almost similar to the method introduced in section 2.2.1.

In the simplification step, quadric is distributed to the leaves of the BSP-Tree containing three vertices of the face. Only the three vertices which can be mapped to three different leaves can be remained. Accordingly, the others degenerate and are discarded. Once one scan is complete, the position of representative vertex of each leaf is computed in second scan by using its associated quadric. The final simplified meshes consist of these vertices and the non-degenerated faces.
This adaptive simplification algorithm is well suited to simplify the large meshes like over-tessellated meshes. Also, this algorithm is memory and time efficient since it makes only two linear scans over the input data and yields a simplification hierarchy which makes it possible to create level-of-detail representations.

2.2.3. Multi-resolution 3D approximations for rendering complex scenes (Rossignac and Borrel, 1993)

This algorithm, which is normally called vertex clustering, operates on boundary representations of an arbitrary polyhedron and generates a series of simplified models with a decreasing number of faces and vertices. Vertex table \( V \) which contains vertex coordinates and a face table \( F \) which contains references to \( V \) represent the original model. The simplification process is summarized in Figure 2-3.

![Figure 2-3: Overview of the simplification process, (Rossignac and Borrel, 1993)](image)

In triangulation, each face is decomposed into triangles supported by original vertices, resulting in table \( T \) which contains three vertex-references per face.

In clustering, the vertices are clustered on the basis of geometric proximity, and stored in two tables: cluster indices \( R \) and clusters \( C \). Table \( R \) indicates the corresponding cluster number and table \( C \) contains a list of references of vertices falling into that cluster.

In grading, the weight of each vertex in table \( V \) is computed and stored in weight table \( W \). High weights are given to vertices which have high possibility of lying on the object’s silhouettes from arbitrary viewing direction and vertices which bound large faces that should not be affected by the removal of small details.
In synthesis, a representative vertex is computed using the table $C$, $W$ and $V$ and stored in table $SV$ of simplified vertices.

In elimination, by comparing the vertices of original triangles and the three new representative vertices in table $R$, table $ST$, $SE$ and $SP$ are created. When all three representative vertices are equal, the triangle degenerates into a point, stored in $SP$ table. When two vertices are equal, the triangle degenerates into edge, stored in $SE$ table. When no vertex is equal, table $ST$ is produced. During this elimination process, $ST$, $SE$, and $SP$ tables contain three, two, and one reference to entries in the $SV$ table.

The increasing degree of simplification can be obtained by executing all the processes several times, with decreasing clustering resolutions. And, a series of models with different degree of simplification can be achieved by toggling between the full resolution of original data and a crude approximation of the data.

With the limitation of hardware graphic performance, this algorithm is said to be well suited for the real-time visualization of complex 3D objects by using the series of increasing simplification techniques which are obtained by storing several simplified representations with different simplification factors in addition to the original model.

2.3. Particle simulation

2.3.1. Re-Tiling polygonal surfaces (Turk, 1992)

This algorithm can be sub-divided into several steps: choosing a set of points, relaxation procedure to calculate the repelling force, connection of the candidate points to form a triangular mesh, old vertices removing, and topology and triangle checking steps.

In the step of choosing a set of points, a user can choose the desired number of vertices and place them randomly in the place which will give well-shaped triangles in the new representation of the surfaces.

Once the desired number of vertices is chosen, the next step is to find the position where those vertices are placed. Basically, relaxation procedure is applied to move each point away from all other nearby points. The repelling force that one point has on another is a force that falls off linearly with distance, and thus becomes zero at a fixed radius. To be more precise, the force vectors can be computed by the following formula (Pauly et al., 2002),

$$F_i(p) = k(r - ||p - p_i||)(p - p_i) \quad (2.3)$$

Where, $F_i(p)$ is the force exerted on the particle $p$ due to the $p_i$, $k$ is a force constant and $r$ is the repulsion radius. The total force exerted on $p$ is given as $F(p) = \sum_{i \in N_p} F_i(p)$

Where, $N_p$ is the neighbourhood of $p$ with radius $r$. The search for nearby points can be made constant-time by placing all points in a three-dimensional grid data structure.
After placing the new vertices on the model’s surface defined by individual polygons, the next step is to connect the candidate points to form a triangular mesh. This algorithm adopts a mutual tessellation that incorporates both the old vertices of the original surface and the new points which will become new vertices.

In simplification step, the old vertices are removed in a way that guarantees the newly-created triangles follow the topology of the original surfaces. Given an old vertex \( R \) to be removed, triangle set \( T \) is defined as a set which share the triangles with vertex \( R \). The set of vertices \( V \) in \( T \) without \( R \) are projected onto the tangent plane of \( R \). And then, a few tests are made to see if this region can be re-tiled without changing the topology of the original surfaces. The tests are done by using several constraints; one constraint is that all edges of the triangles in \( T \) which do not contain \( R \) must be included in the final triangles, and the other constraint is that no new edges are to be introduced outside of the polygon formed by the edges of final triangles. Finally, the topological consistency and triangle shape are checked before removing old vertices.

![Figure 2-4: Removing a vertex(R) from a mutual tessellation, (Turk, 1992)](image)

In Figure 2-4, the result of removing vertex \( R \) and triangulating the neighbours in \( V \) gives five new triangles from the seven triangles, and are constrained to have common borders which are the edges \( AB, BC, CD, DE, EF, FG, \) and \( GA \). So, the newly-created triangles are adjacent to the same triangles that used to be borders of triangles in \( T \).

This algorithm is best suited for the curved surfaces like medical data, molecular graphics, or organic forms such as animals or people with the given number of points, while it is poorly suited for the well defined corners and sharp edges such as buildings since it changes the original position of the points.

### 2.4. Iterative simplification

#### 2.4.1. Point set surfaces (Alexa et al., 2001)

This algorithm tries to remove the points iteratively according to the degree of contribution of a point to the shapes, which incorporate the distance of points from the surface, curvature, and distance from the medial axis of the shape. The actual simplification process can be explained by three steps:
defining the reference plane, computing the coefficients of a polynomial approximation, and generating the representing point set.

In defining reference plane step, the reference plane, \( H = \{ x \mid <n, x> = 0, x \in \mathbb{R}^3, n \in \mathbb{R}^3, \|n\| = 1 \} \) for the point \( r \) is defined to minimize a local weighted sum of squared distances of the point \( p_i \) to the plane. The projection of \( r \) onto \( H \) is assumed to be \( q \), and then \( H \) is found by minimizing \( \sum_{i=1}^{N} (\|n, p_i\| - D)^2 \theta(p_i - q) \). Where, \( \theta \) is a smooth, radial, monotone decreasing function.

![Figure 2-5: The MLS projection procedure, (Alexa et al., 2001)](image)

The next step is to compute the coefficients of a polynomial approximation of \( g \) by minimizing the weighted least squares error \( \sum_{i=1}^{N} (g(x_i, y_i) - f_i)^2 \theta(p_i - q) \). Here \( (x_i, y_i) \) is the representation of \( q_i \) in a local coordinate system in \( H \), and \( f_i \) is the height of \( p_i \) over \( H \), i.e. \( f_i = n \cdot (p_i - q) \). The projection of \( r \) onto \( g \) is determined by the polynomial value at the original, i.e. \( q + g(0, 0) \cdot n \). Here, N-body problem is used to get the coefficient values, and can be found in (Alexa et al., 2003) in detail.

Now, the approximation of polynomial \( g \) can be obtained with the coefficient values, and the representing points set can be generated by removing the unimportant points. In this algorithm, the contribution of a projected point \( p_i \) to the surface \( S_p \) (approximated surfaces) is used as a criterion to determine the importance of that point. This value is estimated by comparing \( S_p \) and \( S_p - [p_i] \). To be more precise, the contribution value of point \( p_i \) is approximated by the distance from \( p_i \) to its projections onto \( S_p - [p_i] \). (Projecting \( p_i \) under the assumption it was not part of \( P \)).

After obtaining the contribution values of all points, these estimated contribution values are inserted into a priority queue, and the point with smallest value is removed iteratively. And then, the values of nearby points are recalculated since they might have been affected by the point removal. This process is repeated until the desired number of points is reached or the contributions of all points exceed some pre-specified bound.
2.4.2. Progressive point set surfaces (Fleishman et al., 2003)

This algorithm decomposes the surface representation of the object into a tangential and a normal component for progressive or multi-scale representations. The projection operator, which maps points in the proximity of the shape onto local polynomial surface approximations, is defined to describe a local tangential coordinate frame which allows us to specify the position of inserted points with a scalar representing the normal component. The base point set is constructed based on the projection operator, and progressive point set surfaces are constructed by using refinement rule which uses local reference domain $H_p(p_i)$ to compute the polynomial fit $g_p(p_i)$ to the input point set.

In Figure 2-6 (a), a new point $a$ is generated in the neighbourhoods of the surface in current point set (the blue points), and the reference domain $H_p(p_i)$ and polynomials $g_p(p_i)$ are computed. A new point $a'$ on $H_p(p_i)$ is generated and projected on the polynomial $g_p(p_i)$.

The simplification scheme is based on the refinement rule to construct a base point set which represents a smoother version of the original shape. In Figure 2-6 (b), given a reference (input) point set $R = \{r_i\}$ defining a reference surface $S_r$, and a base point set $P$ which is derived from $R$ ($P \subset R$) by removing points defines $S_p$. After a point $a$ is projected onto the reference plane $H_p(a)$ and polynomials $g_p(a)$, the point $a$ is projected again onto $S_r$ and $S_p$. The detail value $\Delta = a_r - a_p$ is computed. The criterion for reducing data from input (reference) point set into base point set is that the local reference domain of $p_i$ with respect to $R$ (input point set) and with respect to $P$ (MLS point set or base point set) is about to be same, meaning that $\Delta$ is almost zero.

To be more precise, given a neighbourhood size $h$, let $P_1$ be the reduced point set in h-neighbourhood around $r_i$, e.g. $P_1 = \{|r_i|| r_i - r_i > \h \}$. A point $r_i$ can be used in the base point set if the original reference domain $H_b(r_i)$ is close to the reference domain $H_p(r_j)$ with respect to the reduced point set $P_1$. The distance between $H_b(r_i)$ and $H_p(r_j)$ is measured as the scalar product between their normal components.

This algorithm uses MLS (moving least squares) surfaces to derive the reference domain based on the k-nearest neighbours, inherently contains their properties. However, the lack of connectivity in the
representation is the shortcoming of this algorithm, which requires relatively dense point set to resolve possible ambiguities in the modelling of complex surfaces.

2.4.3. Surface simplification using quadric error metrics (Garland and Heckbert, 1997)

This algorithm, also known as Qslim or vertex merging, is based on the iterative contraction of vertex pairs which are selected according to the cost of contraction. To associate quadric form with the notion of cost of contraction, a symmetric 4x4 matrix $Q$ is introduced. Accordingly, error $\Delta(v)$ at a vertex $v = (v_x, v_y, v_z, 1)^T$ is defined as $\Delta(v) = v^T Q v$. The general simplification process of this algorithm is as follows (Figure 2-7):

1. Compute the $Q$ matrices for all the initial vertices.

2. Select valid pairs. The valid pairs $(v_1, v_2)$ should be an edge or meet the condition which is $\|v_1 - v_2\| < t$ (threshold).

3. Compute the contraction target $\tilde{v}$ for each valid pair $(v_1, v_2)$. The error of this target becomes the cost of contraction of that pair $\Delta \tilde{v} = \tilde{v}^T (Q_1 + Q_2) \cdot \tilde{v}$.

4. Place all the pairs with the minimum cost at the top.

5. Iteratively remove the pair $(v_1, v_2)$ with the least cost placed on the top, contract this pair, and update the costs of all valid pairs involving $v_1$ and $v_2$.

![Figure 2-7: Edge contraction, (Garland and Heckbert, 1997)](image)

To execute this algorithm, the remaining thing is how to calculate the 4 $\times$ 4 symmetric matrices $Q$. Normally, each vertex is the solution of the intersection of a set of planes. In other words, a set of planes meet at that vertex. So, the error of the vertex with respect to this set is defined as the sum of squared distances to its planes:

$$\Delta(v) = \Delta([v_x, v_y, v_z, 1]^T) = \sum_{p \in \text{planes}(v)} (p^T v)^2$$

(2.4)

Where $p = [a, b, c, d]^T$ represents the plane defined by equation $ax + by + cz + d = 0$, where $a^2 + b^2 + c^2 = 1$. Then, the error metric can be rewritten as a quadric form:
\[\Delta v = \sum_{p \in \text{plane}(v)} (v^T p)(p^T v)\]
\[= \sum_{p \in \text{plane}(v)} v^T (pp^T) v\]
\[= v^T (\sum_{p \in \text{plane}(v)} K_p) v\]

Where, \(K_p\) is the matrix:
\[
\begin{bmatrix}
a^2 & ab & ac & ad \\
ab & b^2 & bc & bd \\
ac & bc & c^2 & cd \\
ad & bd & cd & d^2
\end{bmatrix}
\]

Here, \(K_p\) is called fundamental error quadric. Finally, the error quadric \(Q\) for the vertex is the sum of fundamental error quadrics.

While this algorithm shows the combination of efficiency, quality, and generality, there are also several limitations. First, measuring error as a distance to a set of planes only works well in a suitable local neighbourhood. Second, information accumulated in the quadrics is essentially implicit, so it is sometimes problematic when we join two faces since it is not clear to determine which face is defunct.

2.4.4. Point cloud representation (Linsen, 2001)

This algorithm presents very important concepts of defining neighbourhood and the entropy of a point. The simplification is performed gradually based on the entropy of point which is the information content of that point. Here, the concept of neighbourhoods and the criteria of determining the entropy of a point will be explained.

In this method, the connectivity of triangular meshes in faces or surfaces is substituted by the environment of a point, which is the concept of neighbourhoods, since point based method can not contain any faces or surfaces. The author incorporated an angle criterion to define neighbourhood in addition to using k-nearest neighbours, since k-nearest neighbours may possibly cover only the half of the environment as can be seen in the case of Figure 2-8 (a). The introduced method computes the least squares best fitting plane \(P\) of a point \(p\) and its neighbours, and the neighbours are projected onto
P, then angles are made with every neighbour $q_i$ and $q_{i+1}$. This angle $\angle q_i q p q_{i+1}$ must fulfill the angle criterion.

For data reduction, this algorithm uses five criteria which are the distance from a point to its neighbours $\{M_d(p)\}$, non-planarity $\{M_p(p)\}$ which is the distance of the neighbours from the tangent plane at $p$, change of the normal $\{M_c(p)\}$, the non-uniformity $\{M_u(p)\}$, and the colour information $\{M_{co}(p)\}$ which is containing RGB-values of $p$ and its neighbours $p_j$ in $c$ and $c_j$. All of them are defined as follows, where $d_i = \|d_i\|$, $d = \frac{1}{k} \sum_{j=1}^{k} d_j^2$:

$$M_d(p) = \sum_{j=1}^{k} d_j^2$$

$$M_p(p) = \sum_{j=1}^{k} (n'd_j)^2$$

$$M_c(p) = \sum_{j=1}^{k} \|n - n_j\|_2^2 \cdot d$$

$$M_u(p) = \sum_{j=1}^{k} \|(n + n_j)'d_j\|^2$$

$$M_{co}(p) = \sum_{j=1}^{k} \|c - c_j\|_2^2 \cdot d$$

Finally, $M(p)$ is a weighted sum of its components,

$$M(p) = a_0M_d(p) + a_0M_p(p) + a_0M_c(p) + a_0M_u(p) + a_0M_{co}(p) \quad (2.5)$$

The simplification is done on the basis of calculated sum of $M(p)$. However, the way of giving weights has not been explained in details in the paper, but will be different according to the applications.

Figure 2-10 shows two examples of reduced point clouds by considering only distance in Figure 2-10(b) or by considering surface features in Figure 2-10(c) as well. The fine details are lost in Figure 2-10(b), in particular at the teeth.
2.4.5. Meshfree thinning of 3D point cloud (Dyn and Iske, 2007)

This algorithm adopts the concept of significance of every single point as a criterion to reduce the point density. The significance value of a point is derived by computing the local surface approximation which is derived from the radial basis function, or more generally, kernel-based approximation method instead of moving least squares surfaces. To be more precise, the criterion to reduce the point from the original model can be explained by local approximation. The details of that method can be referred to (Buhmann, 2003).

The actual significance value of a point can be calculated by following formula:

\[
\sigma_\infty (x) = \max_{y \in M^x_S} | s(\pi^y_x (y); N^y_x) - \delta (y; N^y_x) |
\]  

(2.6)

Where, \( \sigma_\infty (x) \): Significance value

\( M^x_S \): Test set, being maintained during the thinning process, This considers the previously removed points as well as currently remaining points

\( S \): The local approximation derived from kernel-based approximation

\( N^y_x \): The neighbouring point set of \( x \)

\( \pi^y_x (y) \): The projected point set of local neighbourhood \( N^y_x \) onto the tangent plane

\( s(\pi^y_x (y); N^y_x) \): The distance of projected point set to the local approximation

\( \delta (y; N^y_x) \): The signed distance of \( y \) to the direction of neighbourhood.
Figure 2-11: Illustration of procedures

Figure 2-11 shows the simplified procedures of calculating significance value of point $x$. The neighboured point $y_1$ of $x$ is projected to the tangent plane as $\pi^T(x)$ (blue point); $\delta(y; N^T_x)$ is calculated in the direction of $y_2$; $s(\pi^T(x); N^T_x)$ is calculated by projecting $\pi^T(x)$ to the local approximation(s) of neighbouring points set $(N^T_x)$. Finally, the significance value $\sigma_\infty(x)$ becomes the maximum difference between $s(\pi^T(x); N^T_x)$ and $\delta(y; N^T_x)$ (the length of blue line).

The thinning is done by removing the point with least significance value repeatedly. After removing the point $x_i$ with least significance value, the calculation of significance value of a certain point is done by using same neighbourhood sets as it used when removing point $x_i$. To be more precise, the neighboured points of a certain point to be used for calculating significance value are being maintained during thinning process. The real thinning is accomplished by repeating the process of removing the least significant point $x_i$ and updating the significance value of remaining points with the same test set repeatedly.

Figure 2-12: Result of data reduction, (Dyn and Iske, 2007)
As can be seen in the results of reduced data set with Dragon in Figure 2-12, this algorithm seems to be able to maintain geometric features and their topology correctly in the continuous data set without big gaps.

2.5. Summary and requirements for a new method

2.5.1. Analysis on the introduced literatures

The summary of methods mentioned above is given in Table 2-1. This includes the criteria which have been used to reduce the data in each method and several characteristics of them.

<table>
<thead>
<tr>
<th>Category</th>
<th>Algorithm</th>
<th>Criteria for reduction</th>
<th>Characteristics</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering</td>
<td>Model simplification through refinement (Brodsky and Watson, 2000)</td>
<td>Face normal variation $nv = \sum_{i=1}^{N} a_i (1 - cp)$</td>
<td>• Well suited for large model</td>
<td>Called as R-simp</td>
</tr>
<tr>
<td></td>
<td>Efficient adaptive simplification of massive meshes (Shaffer and Garland, 2001)</td>
<td>Quadric - squared distance from a vertex to a set of planes Dual quadric - squared distance from a plane to a set points</td>
<td>• Suitable for over-tessellated meshes • Memory and time efficient Covariance matrix in PCA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-resolution 3D approximations for rendering complex scenes (Rossignac and Borrel, 1993)</td>
<td>Clustering Grading Triangulation</td>
<td>• Suitable for multiple degree of simplification</td>
<td></td>
</tr>
<tr>
<td>Particle simulation</td>
<td>Re-Tiling polygonal surfaces (Turk, 1992)</td>
<td>Repelling force - the force that one point falls off linearly to another with distance</td>
<td>• Suitable for curved surfaces such as animals or people</td>
<td></td>
</tr>
</tbody>
</table>
## Table 2-1: Summary of the algorithms for data reduction

<table>
<thead>
<tr>
<th>Iterative simplification</th>
<th>Contribution of a point to the shape (estimated by distance between projected point $p_i$ to the surface $S_p$)</th>
<th>Moving Least Squares surfaces (MLS surface)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressive point set surfaces (Fleishman et al., 2003)</td>
<td>Difference between two local reference domains of $p_i$ with respect to $R$ and in $P$ ($R$ : input point set) ($P$ : base point set)</td>
<td>K-nearest neighbourhood MLS surface</td>
</tr>
<tr>
<td>Surface simplification using quadric error metrics (Garland and Heckbert, 1997)</td>
<td>Contraction cost of vertex pair (squared distances from vertex pair to its planes)</td>
<td>Optimal combination of efficiency, quality, and generality called as Qslim / Vertex merging</td>
</tr>
<tr>
<td>Point cloud representation (Linsen, 2001)</td>
<td>Surface features: $M(p)$ $\alpha_0 M_d(p)+\alpha_0 M_n(p)+\alpha_0 M_c(p)+\alpha_0 M_u(p)$</td>
<td>K-nearest neighbourhood with angle criteria</td>
</tr>
<tr>
<td>Meshfree thinning of 3D point cloud (Dyn and Iske, 2007)</td>
<td>Significance value</td>
<td>Local surface approximation</td>
</tr>
</tbody>
</table>

According to the literatures introduced here, several important conclusions can be derived for developing a new method.

First, curvature information of the surfaces was adopted in many methods in certain forms; which are the covariance matrix of face normal variation in (Brodsky and Watson, 2000), Quadric form in (Shaffer and Garland, 2001), and non-planarity $M_p(p)$, change of normal $M_c(p)$ and the non-uniformity $M_u(p)$ in (Linsen, 2001). Basically, the points on the curved areas have higher values based on the curvature information by those methods; accordingly that higher value can guide the algorithm to select more samples from the curved area than the flat area.

Second, distances between one point and some other targets were considered in many methods. They were expressed in many forms: proximity in (Rossignac and Borrel, 1993), distance from a point to its
projection onto local MLS surface in (Alexa et al., 2001), the distance between the two reference domains in (Garland and Heckbert, 1997), the distance from a point to its neighbours $M_d(p)$ in (Linsen, 2001), maximum difference between the distance of local surface approximation to the projected neighbouring point, and signed distance of one neighbouring point to the other neighboured point direction (Dyn and Iske, 2007). Such distances always play a key role to determine the importance of that point. In that sense, the point far from the other point was regarded as more important point than the points near together. That is certainly because if two points are very close, one of them can have higher potential to be a redundancy.

Third, locality has been a quite important issue when calculating value of a point, so the methods by (Alexa et al., 2001), (Fleishman et al., 2003), and (Dyn and Iske, 2007) define the local reference domain or local surface approximation based on the pre-defined neighbourhoods. They define their own neighbourhoods when calculating some values of the points. In other words, the importance of a certain point has been reflected by their local surroundings. Hence, the value of point can be affected a lot by the way how the local surroundings are defined. So, special cares should be given when defining locality.

### 2.5.2. Requirements for a new method

From the consequences derived above, several important ideas for designing a new method can be derived. The new method which will use significance value as a criterion to remove certain points should meet the following conditions;

i. The significance value of a point should be able to reflect the properties of local surroundings
   - The significance value should incorporate the properties of curvature and distance information of their surroundings
   - In the case of buildings, the points on the edges should have higher significance value than the points on the planar or flat area.
   - In the case of terrain, the points on the break line should have higher significance value than the points on the regularly sloped area, and the points on the region of heavy changes in curvatures should have higher value than a planar region or a region of constant curvature.

ii. The range of significance value should be categorized into certain classes based on the properties of the surface.

iii. The reconstructed surfaces from the reduced data set should have similar surfaces constructed from the original data set.
3. Proposed method for data reduction

This research tries to develop a new point-based simplification method being able to remove the points which contribute less to represent the surfaces of the earth. In the proposed method, every single point is evaluated whether to be remained or removed based on the demands of points to represent the surfaces, which is called significance value in this paper. If the point has a high potential to be used in the process of reconstructing surfaces, the significance value of this point will be high. The formulation of criteria calculating significance is a core of this research. The criteria are embodied as two parameters. Besides, algorithm to compute the actual significance value of any point in a given point cloud and given parameters was developed. After computing significance values, points are selected by giving thresholds on calculated values. Here, the way to find the threshold values by each parameter will be proposed. Finally, the optimal combinations of two threshold values for the specific surface properties will be proposed.

This chapter describes the method how the significance value of each point has been calculated. The required pre-processing to derive two parameters which are sum of height differences and change of normal within the neighboured points is explained in section 3.1. Defining the neighboured points is done in 2D space (Fix and Hodges, 1951) and the normal vectors are derived by applying the plane fitting method (Hoppe et al., 1992). After pre-processing, the methods used for calculating significance value of points are described separately in section 3.2. The overall workflow used in this research is as follows;

![Figure 3-1: Workflow applied in the research](image-url)
3.1. Pre-processing

The information to be used in this research for calculating significance value of each point are 3 dimensional coordinates and normal vector of point with respect to the local surfaces defined by neighboured points. 3 dimensional coordinates of points can be obtained directly from the airborne laser scanning data; however, the normal vector of point should be calculated in a certain way. So, in this section, the way of defining neighbourhoods will be described at first, and then the method of calculating the normal vector of every point with respect to the locally defined surfaces will be described.

3.1.1. Defining the neighbourhood

To fulfil the first requirement of a new method, which was mentioned as the significance value should be able to reflect the properties of local surroundings, the output by applying the proposed algorithm should be interpreted as point of higher value is more significant than the point of low value. Here, the information to be used when calculating significance values is obtained from the pre-defined local surroundings. Hence, proper definition of the local surroundings is one of the most important procedures of this research.

In general, local surroundings can be defined in two ways: in 2 dimensions or in 3 dimensions. In this research, k-nearest neighbourhoods in 2 dimensions are adopted as a local surrounding. The reason that 2 dimensions are selected is that if we look at the local surroundings in 3 dimensions, the neighboured points can not reflect the properties of real surroundings of that point. For example, for the seed point (red point) in Figure 3-2, the neighboured points in 3 dimensions will only be selected from the roof of the objects appearing in blue points, while the neighboured points in 2 dimensions can be selected from the grounds as well as from the roof of the buildings, appearing in blue or green points. So, the seed point (red) could be better reflected by these 2 dimensional neighboured points.

For more sophisticated definition of neighbourhood, Floater’s method based on local Delaunay triangulations(Floater and Reimers, 2001) or Linen’s angle criterion(Linsen, 2001) method could be referred to.

![Figure 3-2: Neighbourhood in 2 dimensions and 3 dimensions](image-url)
3.1.2. Estimation of normal vector

The other information needed for further processing is the normal vector of all points with respect to surfaces defined by neighboured points. Basically, the normal vectors can be defined from the surface as seen in Figure 3-3. However, the problem is how to estimate the normal of the surface at that point. Two main approaches are described in the paper by (Dey et al., 2005): one is numerical approach applying some optimization technique, the other is combinatorial applying some Delaunay property. In this research, the plane fitting method (Hoppe et al., 1992) which in one of the widely used techniques in numerical approach was adopted.

![Figure 3-3: A normal to a surface or tangent plane at point \( p_i \)](image)

To estimate the oriented tangent plane \( T(p_i) \) of the point \( p_i \), the pre-defined neighboured points \( N(p_i) \) are. So, the \( T(p_i) \) is defined as follows;

\[
T(P_i) : (p_{ik} - p_i) \cdot v_0 = 0 \tag{3.1}
\]

Where, is \( p_i \) is the seed point which we want to get a normal vector within neighbour points \( N(p_i) \) which is the minimized sum of squared distance from \( p_i \) to \( N(p_i) \), and \( v_0 \) is the approximated normal of surface at the point \( p_i \). The normal of the point can be determined by using principal component analysis (PCA)(Jolliffe, 2002). To compute the \( v_0 \), the 3x3 covariance matrix \( C \) with \( N(p_i) \) is formed as follows;

\[
C = \begin{bmatrix}
    p_{i1} - p_i & ... & p_{i3} - p_i \\
    ... & ... & ... \\
    p_{ik} - p_i & ... & p_{ij} - p_i \\
\end{bmatrix} \cdot \begin{bmatrix}
    p_{i1} - p_i \\
    ... \\
    p_{ik} - p_i \\
\end{bmatrix}, p_{ik} \in N(p_i) \tag{3.2}
\]

Where, \( p_i \) is the seed point which we want to get a normal vector within neighboured points \( p_{ik} \).

From the eigenanalysis of covariance matrix \( C \) within a local neighbourhood \( N(p_i) \) which meet the \( C \cdot v_0 = \lambda_i \cdot v_0 \), \( \lambda \in \{0,1,2\} \) , the tangent plane \( T(p_i) \) and normal vector of the point \( p_i \) can be determined by analyzing the values from \( \sum_{p_{ik} \in N(p_i)} |p_{ik} - p_i|^2 = \lambda_0 + \lambda_1 + \lambda_2 \). Assuming \( \lambda_0 \leq \lambda_1 \leq \lambda_2 \), \( v_0 \) approximates the surface normal at point \( p_i \) as a form of unit normal vector, with \( v_1 \) and \( v_2 \).
spanning the tangent plane at \( p_i \) as was explained in section 2.2.1 (Brodsky and Watson, 2000). In addition, \( \lambda_o \) can quantitatively estimate how much the point deviate from the tangent plane, while \( \lambda_1 \) and \( \lambda_2 \) describe the variation along point distribution in tangent plane showing estimation of local anisotropy. \( \lambda_o = 0 \) can be interpreted as all points lie in the plane. In the research, only the normal unit vector \( V_o \) which has the length of 1 is extracted for further processing.

However, strictly speaking, the unit normal vector field for the tangent plane \( T(p_i) \) can be defined exactly in two ways pointing in opposite directions \( (V_o \) or \( -V_o) \), provided that the unit normal vector field can be defined on all of surfaces at the point \( p_i \). So, to get a consistent orientation of unit normal vectors, the method based on the minimum spanning tree (Hoppe et al., 1992) is adopted in this research. The basic idea of this method is that the normal vectors of two close tangent planes should be almost parallel and should have a consistent orientation, if it is not; one normal vector is flipped by changing their sings. By adopting this method, consistent calculation of change of normal can be obtained.

### 3.2. Derivation of parameters for calculating significance value

In this section, two different parameters which are sum of height differences and changes of normal vector within neighboured points will be introduced. In addition, the way of selecting significant points by applying threshold values (T1, T2) will be explained for practical applications.

#### 3.2.1. Calculating sum of height difference within neighboured points

Laser scanning data contains 3 dimensional coordinates and some other scanning information. Among them, the \( x-y \) coordinates are closely related to the point density and scanning directions, while \( z \) coordinate is reflecting the height information of that point. One parameter to calculate significance value of the point uses these 3 dimensional coordinates. Within the pre-defined neighboured points, the \( z \) coordinates may change a lot in irregular surfaces. So, if the height differences between one point \( P_i \) and its neighboured points are summed up, the results will be different according to the properties of surfaces.

\[
H(\ P_i\ ) = \left( \sum_{k=1}^{n} (P_k(z) - P_i(z)) \right), \ i = 1, 2, 3...m. \tag{3.3}
\]

Where, \( H(P) \) = sum of height differences between neighboured points

- \( P_k(z) \) = \( z \) coordinate of neighboured points
- \( P_i(z) \) = \( z \) coordinate of point \( P_i \)
- \( n \) = number of neighboured points
- \( m \) = number of point in the data set.

From this formula, if the seed point is located in the planar area, the summed value will become almost zero regardless of the slopes or aspects since the height differences will be compensated by the point located in the opposite side of that point (in Figure 3-4(a)). On the other hand, the sum of height differences between seed point and neighboured points in the curved surfaces may not become zero, but resulting in some positive or negative value (in Figure 3-4(b)). In other words, this value can
represent the planarity of the surfaces very well since the non-zero value can be interpreted as a seed point is not located on the planar surfaces.

![Diagram showing points on planar and curved surfaces](image)

**Figure 3-4: Sum of height differences**

To get a compatible value of point as an indicator of representing properties of local surroundings of the points, several steps of modifications are required. For example, the points located on the border of high objects like a building roof will have a large negative value due to the big height differences caused by neighboured points on the grounds. Furthermore, not only the border points but also some points which have any ground point as their neighbours will have very large values also, which makes it difficult to distinguish relative significances among them. That is why some steps of modifications are required.

The first step of modifications is to take an absolute of the calculated value in order to convert them into positive values. In reality, the sign of value (+ or –) only indicates whether the seed point is located relatively below the neighboured points or above the neighboured points. The sign of the value has nothing to do with the significance of the point. Only the values without sign can be interpreted as an indicator of significance of the point. For example, both the points on the break lines like edges or ridges of objects, and points on the valley will have high values similarly. On the other hand, the points on the planar areas or in the middle of symmetric surroundings will have small values. So, the formula becomes:

\[
H (p_i) = \text{abs} \left( \sum_{k=1}^{n} (P_k(z) - P_i(z)) \right), i = 1, 2, 3...m. \tag{3.4}
\]

The second step of modification is to incorporate the maximum height differences between two points in order to prevent one specific point from affecting the value too much. For example, for the points located close to the border of tall objects like in Figure 3-5, one or two points from the grounds may be included as its neighboured points. In that case, the calculated value of that point will be very large due to the big height difference between seed point and a few ground points. However, from the view
point of significance, the point located inner side of the borders may not be significant point, even though the calculated value is high.

![Figure 3-5: necessity for incorporating maximum height differences](image)

So, the height differences \((P_k(z) - P_i(z))\) exceeding a certain value should be set to have a certain value as a maximum. In the proposed algorithm, maximum height difference is constrained to have a value of three times of distance between two points as a maximum value. Actually in the real world, the roof slope in normal buildings is less than 70°, which is equivalent to slope 3, except for some religious buildings or towers. So, the maximum height difference, which is the 3 times of distance between two points, can be said to be reasonable in general purposes. So, the formula is modified as follows;

\[
H(p_i) = \text{abs} \left( \sum_{k=1}^{n} \text{Min} \left( 3 \times \text{dist}(P_k, P_i), \left( P_k(z) - P_i(z) \right) \right), i = 1, 2, 3 \ldots m. \right) \tag{3.5}
\]

Where, \( \text{dist}(P_k, P_i) = \sqrt{(P_k(x) - P_i(x))^2 + (P_k(y) - P_i(y))^2} \)

From the formula (3.5), if the height difference between two points \((P_k(z) - P_i(z))\) is 1.0 m and the distance is 0.1 m, the resulting height difference will become 0.3 m (0.1 m x 3). In other words, all slopes larger than 3 (larger than 70°) will be treated uniformly.

![Figure 3-6: Maximum height difference](image)

Another issue when using height differences as a parameter is that the values will be changeable according to the resolution of the data. To be more concrete, the values from low resolution data may be higher than the values from high resolution data. That is because distance and height difference between two points in low resolution data will be larger. So, the further modification is required to set the values to be resolution-invariant, thus taking the slopes between seed point and neighboured points instead of taking only height differences between them as input values. Finally, the formula is modified as follows;

\[
H(p_i) = \text{abs} \left( \sum_{k=1}^{n} \text{Min} \left( 3 \times \frac{P_k(z) - P_i(z)}{\text{dist}(P_k, P_i)} \right), i = 1, 2, 3 \ldots m. \right) \tag{3.6}
\]
3.2.2. Calculating change of normal vectors within neighboured points

After pre-processing, every point has its own unit normal vector with respect to their tangent plane. So, the change of normal vectors between two points can be calculated by subtracting each other, resulting in other vector. As can be seen in Figure 3-7, the maximum difference between two points will occur when two unit normal vectors face to the opposite directions. In that case, the length of resulting vector will become 2. In other words, the amount of changes of normal vectors between two points will be expressed as absolute value from 0 to 2. So, the sum of those absolute values will become a total amount of normal changes \( Nc(P_i) \) within neighboured points, being expressed as:

\[
Nc(P_i) = \sum_{i=1}^{n} \| P_k(N) - P_i(N) \| , \quad i = 1, 2, 3...m. \quad (3.7)
\]

Where, \( P_k(N) \) is a unit normal vector of a point \( P_k \)

\( n \) is a number of neighboured points

\( m \) is a number of points in data set

However, the summed value of normal changes will be different according to the number of neighboured points, which requires normalization by the number of neighboured points in order to get a compatible value regardless of the number of neighboured points. So, the calculated values for all points can be set to have the range from 0 to 2 by following formula:

\[
Nc(P_i) = \sum_{i=1}^{n} \frac{\| P_k(N) - P_i(N) \|}{n} , \quad i = 1, 2, 3...m. \quad (3.8)
\]

From formula (3.8), 0 can have a meaning that the overall changes of directions of normal vectors do not happen, while 2 has a meaning that the changes of directions of normal vectors show the maximum. However, this value does not seem to be easily understandable by users, instead change of normal vectors can be better understandable by means of degrees rather than simple numbers. So, the values are converted into degrees by applying the law of cosines \( P_i(N) \cdot P_j(N) = \| P_i(N) \| \| P_j(N) \| \cos P_i \), resulting in the following formula (3.9):
\[ \text{ang} \ (P_i) = \cos^{-1}(1 - \frac{(Nc(P_i))^2}{2}) \] (3.9)

Figure 3-8: Angle between two unit normal vectors

3.3. Data reduction

After calculating the values of all points by two parameters, data reduction can be done by selecting some parts from original data by applying one parameter or two parameters together. However, the problem is how to combine two different parameters properly since each parameter has its own pros and cons. For example, the sum of height differences can be a better parameter for the data containing break lines like borders or ridges, while it may not be a good parameter to the inflection points or in the fluctuating area. In other words, the selection of parameter to be applied will be different according to the properties of the data set.

Figure 3-9: Methodology of data reduction

In the proposed method, the data reduction is done in two ways. One way is to give the threshold values by each parameter separately, and then combine them. Any point which has a higher value by one of two parameters will be remained by the proposed method in Figure 3-9. By doing so, the risk of removing significant points only by one parameter can be decreased. The general methodology for data reduction is presented in Figure 3-9, where the ranges of values when selecting two thresholds...
(T1, T2) will be based on the following Table 3-1. The visualization of calculated values can be useful when selecting threshold values, and will be explained in section 3.4.1 in detail.

<table>
<thead>
<tr>
<th>Number of neighboured points</th>
<th>Sum of height difference (T1)</th>
<th>Change of normal (T2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0 ~ 15</td>
<td>0 ~ 180</td>
</tr>
<tr>
<td>20</td>
<td>0 ~ 60</td>
<td>0 ~ 180</td>
</tr>
</tbody>
</table>

Table 3-1: The range of initially calculated values by two parameters

The other method is to use the classification tables which will be proposed later in this research as a form of Table 3-2. The classes of optimal combinations of two threshold values based on the remained rate and RMSE will be proposed with the descriptions of possible surface properties. Here, to get RMSE, which is the difference between DEM generated from the original data and DEM from the reduced data, accuracy assessment will be performed, and will be explained in section 3.4.2 in detail. In addition, the optimal threshold values within the same remained rate will be selected from which has the smallest RMSE. Actually, this method can be said to include first method since all threshold values should be selected before making this table.

<table>
<thead>
<tr>
<th>Class</th>
<th>Remained rate (%)</th>
<th>RMSE (m)</th>
<th>T1 (m)</th>
<th>T2 (°)</th>
</tr>
</thead>
</table>

Table 3-2: The table of classification

3.4. How to find threshold values

3.4.1. Visualization of calculated values

One way of data reduction is to designate the ranges of threshold values of two parameters based on user’s interests. However, the problem is how users can easily recognize the relative significance of each point according to the calculated values. Moreover, the values may not be ranged over whole ranges of the values, which make it difficult for users to distinguish the relative significances of the points. So, the adopted technique in this research to help users is visualization of equalized values in colour space by using histogram equalization method.

Histogram equalization is normally used in image processing. It uses the intensity values and population of them to distribute population evenly over whole ranges as can be seen in Figure 3-10.
In this research, the calculated values by each parameter are used instead of intensity value, and visualized by hue from colour red to cyan in HSL colour space. From this colour space, a user can specify colours intuitively making them a good choice in user interfaces. And then, convert this selection in HSL colour space to RGB colour space by using conversion method by (Gonzalez and Woods, 2007) to visualize them. In this research, the corresponding angles and colours will be set as 0º to cyan and 180 º to red colour based on the strength of impression to human eyes. After selecting certain ranges from 0 and 180, these selected ranges should be converted into original values to be applied in data reduction algorithm. The original threshold values can be obtained easily by selecting the same ranking values in original ranges corresponding to the equalized values.

### 3.4.2. Differences of DEM by original and reduced data

#### 3.4.2.1. RMSE as a measure of accuracy assessment

Normally, the surface of terrain is represented as a DEM which can be a raster or vector format. In this research, DEM as a raster image which is composed of regular grid cells is used. Basically, each grid cell in DEM is represented by one value thus can be statistically analyzed by comparing them. For this purpose, every grid cell from DEMr and DEMo will be subtracted within the same boundary; thus the quality of reduced point set can be assessed quantitatively. Here, DEMo is regarded as reference data (true value) in this case.

\[
\text{Quality} = \text{DEM}_r - \text{DEM}_o
\]  

(3.10)

Where, DEMr is a DEM from reduced data set and DEMo is a DEM from original data set.

From the equation (3.10), Quality is also a raster image which is composed of cells with location and height differences information. In ideal case, the mean and standard deviations of height difference may become zero, resulting in Quality = 0 in every cell. However, the mean value can not be used as a standard to evaluate the quality of reduced point set since the mean value can become zero if the big positive and negative differences are compensated each other. Also, we are not interested in their signs; instead we are interested in the amount of difference. In general, RMSE can be the most
appropriate indicator when we talk about the accuracy of a certain model or estimator compared to the reference data. In that sense, the \( \text{RMSE} \) (root mean square error) from two DEMs’ subtraction is adopted as a standard in terms of \textit{Quality} in this research. By doing so, the small value of \( \text{RMSE} \) will have a meaning that the overall difference between two DEMs is small.

\subsection*{3.4.2.2. Subtracting one DEM from the other DEM}

Prior to DEM subtraction, two DEMs should be generated at first. TIN (Triangulated irregular network) is generated from the points as an intermediate step, and then two DEMs are generated by applying a certain interpolating technique.

![Image of DEM generation](image)

(a) Point data                                    (b) TIN                                             (c) DEM

Figure 3-11: Procedure of DEM generation

One thing to consider here is how to select the appropriate cell size when generating DEMs. Basically, the more accurate surface representation can be obtained with the smaller cell size. However, if too small cell size is chosen, it will definitely require huge computational and memory demands. So, the selection of cell size can be regarded as the tradeoffs between accuracies and efficiencies. In this research, the cell size is selected to be less than the minimum point spacing of original data, which is 0.05 m. By doing so, no two points will be included within the same cell, otherwise might lead to big errors in such an area where one point on the roof and one point from grounds are taken within the same cell.

![Image of cell size selection](image)

Figure 3-12: selection of cell size

After generating two DEMs in this way, the common area of DEM is cropped by multiplying the same type of raster image to each DEM as shown in Figure 3-13. Now, derived DEM1 and DEM2 have same extent of boundaries and can be subtracted.
3.4.3. **Visual check**

After comparing the differences between two DEMs in terms of RMSE, visual checking for the remained points is done to analyze which properties of surfaces are reflected with certain remained rate and RMSE. The description of surface properties for the remained points will be added to those combinations together with remained rate and RMSE to fill in the proposed classification table.

3.5. **Accuracy assessment**

Basically, the significance values were defined in terms of significance to the surfaces, which means that the comparison between reconstructed surfaces and original surfaces should be a main factor for evaluation. In that sense, the comparison of DEMs differences which was used in the second method is the meaningful way for accuracy assessment.

3.6. **Conclusion of the proposed method**

3.6.1. **Summary**

The proposed method incorporated the important concepts derived from introduced literatures in chapter 2 into two parameters. Sum of height difference together with change of normal vectors can reflect the important concepts like distances, curvatures, and locality matters. Sum of height differences is calculated by using the slopes as an input value with constraint of maximum 3 (70º). And change of normal vector is calculated by using unit normal vector of a point which is obtained from eigenvector analysis.

With two derived parameters, two data reduction methods are proposed; one way is to specify the ranges on the visualized values in the colour space, and the other way is to use the proposed classification table. Threshold values are used in both ways to remove insignificant points.
3.6.2. Comparison with other methods

Point based simplification methods were categorized into three classes in chapter 2, which are clustering, particle simulation, and iterative simplification method. Most of point based simplification methods used to be categorized into one of these classes. However, the proposed method cannot be categorized into any of these classes. It seems similar to the iterative simplification method, but it does not calculate the values of point iteratively. It calculates the values of all points at one scan, and applies the reduction method directly. It does not cluster, simulate, and iterate either.

Another characteristic of the proposed method is that a user can interactively select the desired points based on his interests. The currently existing methods only could remove the points in their predefined ways. A user is not involved in the reduction procedures interactively. However, in the proposed method, a user can specify any ranges of calculated values in the first method; also, he can adjust any of two threshold values in the selected class if he is more interested in the specific properties in the second method. The provided results of accuracy assessment will be very helpful when adjusting threshold values.
4. Implementation and results

This chapter describes study areas, selected data set, the implementation of the strategy designed in chapter 3, and the results obtained by applying the proposed method.

4.1. Study area and input data

The study area of this research is small part of Enschede in the Netherlands, and the data set was acquired by FLI-MAP 400 system boarded on the helicopter in March 2007. FLI-MAP 400 system consists of airborne laser scanner, a digital aerial, and two video cameras. This system produces point clouds and orthoimages. Point density depends on how many strips are overlapped and how many returns per pulse are recorded.

The point clouds used in this research contains 6 scanning strips with point density of around 10 points per square meter (10 pts/m²) and flight height was 275 meters. The systematic errors (offsets between strips) are 4-8 cm in X-Y coordinates and 2-3 cm for Z-coordinates. The stochastic error is about 2-3 cm for X, Y and Z coordinates.

4.2. Calculating significance value

4.2.1. Test with a building

The building used in the test is the town hall of Enschede which is composed of several roofs with different heights and a tower as can be seen in Figure 4-1.

![Figure 4-1: Town hall of Enschede](image)

(a) Oblique view  (b) Top view (colours according to the heights)

4.2.1.1. Significance value by sum of height difference

In this section, the values of points by applying sum of height differences within neighboured points will be presented. According to equation (3.6), the value will increase if the number of neighboured points is increased. In the test, the values are calculated with two different numbers of neighboured
points: 5 neighboured points and 20 neighboured points. The calculated values show the ranges from 0 to 15 with 5 neighboured points and from 0 to 60 with 20 neighboured points, showing the distributions in Figure 4-2.

![Figure 4-2: Distribution of sum of height difference](image)

The calculated values are visualized in Figure 4-3 (a) and Figure 4-3 (b) with 5 neighboured points, in Figure 4-3 (c) and Figure 4-3(d) with 20 neighboured points respectively. In addition, histogram
equalization is applied in Figure 4-3(b) and Figure 4-3(d). Here, points with large values are expressed in red colour, while the points with small values are expressed in cyan colour.

4.2.1.2. Significance value by change of normal

Now, the significance values are calculated by applying equation (3.9) as was explained in section 3.2.2. The calculated values of points in building show the ranges from 0 to 180 degree regardless of the number of points, and are visualized in the same way to by sum of height difference.
4.2.1.3. Analysis on significance values of the building

From histogram and visual comparisons of calculated values from Figure 4-2 to Figure 4-5, the calculated values of the points are large in the areas like on the border of the roof or other objects, intersections between buildings and the ground, vegetation, outline of windows or suddenly changing parts on the roof, and etc. They appear in red colours by both parameters. The points on the ridges have intermediate values, appearing in yellow colours. The reason that the points on the ridges have lower values than the points on the borders is due to the gentle slope of the roof. The points on the planar areas have small values, appearing in greenish or cyan colours. From the above observations, two distinctive lessons can be derived.

First, overall distribution patterns of calculated values appear similarly regardless of the number of neighboured points by both parameters. However, strictly speaking, the values by 5 neighboured points are reflecting the small local changes more sensitively. For example, some changes of colours can be seen on the roof in Figure 4-2(b), but not in Figure 4-2(d). This characteristic appears similarly in the test by change of normal. So, it can be said that the large number of neighboured points are not necessary in the proposed method.

Second, the points on edges of objects are more clearly distinguishable by sum of height difference rather than change of normal. That is because the points on the edges are surrounded by points with big height jumps which contribute to the calculated value a lot. In other words, sum of height difference can be regarded as more appropriate parameter for the discrete objects like building.

4.2.2. Test with terrain

The terrain used in the research is the campus of University Twente in Enschede which is a relatively planar area with several small mounds. These areas can give a good impression how the proposed two parameters can reflect the properties of terrain data.

![University Twente in Enschede](image)

Figure 4-6: University Twente in Enschede
4.2.2.1. Significance value by sum of height difference

The calculated values by sum of height difference in terrain data show the same ranges as in the building. However, the distribution patterns of calculated values are somewhat different from the distribution patterns by building. Most of values are ranging lower due to the large part of planar area in the terrains as can be seen in Figure 4-7.

![Figure 4-7: Distribution of sum of height difference](image)

(a) 5 neighboured points  (b) 20 neighboured points

![Figure 4-8: Visualization of sum of height difference](image)

(a) Original values (5N)  (b) Equalized values (5N)
(c) Original values (20N)  (d) Equalized values (20N)
4.2.2.2. Significance value by change of normal

Significance values by change of normal in terrain data are showing the results in Figure 4-9 and Figure 4-10. The distribution patterns appear almost similarly to the distribution patterns by sum of height difference. However, relatively a little wider distributions of values can be recognized compared to Figure 4-7.

![Figure 4-9: Distribution of change of normal](image)

(a) 5 neighboured points  
(b) 20 neighboured points

Figure 4-9: Distribution of change of normal

![Figure 4-10: Visualization of change of normal](image)

(a) Original values (5N)  
(b) Equalized values (5N)  
(c) Original values (20N)  
(d) Equalized values (20N)

Figure 4-10: Visualization of change of normal
4.2.2.3. Analysis on the significance values of the terrain

From the test with terrain data, the overall results are not so different from the result with building data. The large values appear in the area of break lines like edges of the objects, intersections, vegetation, and also in the surface of curved objects. The values are changing gradually according to the curvatures as can be seen in the right upper parts of Figure 4-10(b). From the significance values by change of normal, the following lessons can be derived.

First, like in the case of building, the number of neighboured points does not affect the results much. However, it is recognized that the values with smaller numbers of neighboured points can reflect the small changes of terrain more sensitively similar to the test with building. For example, the small changes within the left lower part of area can be recognized more clearly in Figure 4-8 (b) and Figure 4-10 (b), but not in Figure 4-8 (d) and Figure 4-10 (d). That is because the values become more averaged by the large numbers of neighboured points.

In addition to that characteristic, the distinctive difference in the resulting values by change of normal is shown in the curved area. The points on the curved area have higher values by change of normal rather than by sum of height difference. For example, most points on the mounds in upper right part of area are coloured in red or yellow by change of normal, while less points are coloured in red or yellow by sum of height difference. From this observation, it can be said that the change of normal can better reflect the properties of curved areas. In other words, the change of normal can be regarded as more appropriate parameter for the terrain.

4.3. The result of data reduction by two parameters

4.3.1. Data reduction

In the test, 5 neighboured points were used since which smaller numbers of neighboured points can reflect the small local changes more sensitively as was explained in previous section. The threshold values are selected based on the visualization of equalized significance values. Basically, 64 combinations of two threshold values for each data set are put into test according to the predefined 8 threshold values for each.

After selecting threshold values, RMSE (RE) are calculated by using the function of spatial analyst in ArcGIS software. Following Table 4-1 and Table 4-2 are showing how the remained rate (in the column %) and RMSE (in the column RE) are affected by gradual changes of two threshold values (T1, T2). Here, T1 is a threshold value by sum of height difference; T2 is a threshold value by change of normal.

4.3.2. Analysis on the result of data reduction and RMSE

From Figure 4-11, Figure 4-12, Table 4-1, and Table 4-2, the threshold values show in reverse proportion to the remained rate, but in proportion to the RMSE in overall. In other words, if a user increases one of two threshold values, the data will be reduced more while the RMSE will be increased. So, the remained rate and RMSE should be checked together when determining threshold values to be applied. For example, from Table 4-1, when a user needs to reduce the data less than 30 %, the best combination of threshold values will be about T1 = 2 & T2 = 45 which is showing the lowest RMSE.
DATA REDUCTION OF POINT CLOUDS ACQUIRED BY AIRBORNE LASER SCANNING

<table>
<thead>
<tr>
<th>T1</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>71.7</td>
<td>67.7</td>
<td>66.7</td>
<td>66.6</td>
<td>66.6</td>
<td>66.6</td>
<td>66.6</td>
<td>66.6</td>
</tr>
<tr>
<td>10</td>
<td>61.8</td>
<td>52.3</td>
<td>48.8</td>
<td>48.6</td>
<td>48.6</td>
<td>48.6</td>
<td>48.6</td>
<td>48.6</td>
</tr>
<tr>
<td>15</td>
<td>60.0</td>
<td>49.3</td>
<td>44.9</td>
<td>44.6</td>
<td>44.6</td>
<td>44.6</td>
<td>44.6</td>
<td>44.6</td>
</tr>
<tr>
<td>30</td>
<td>54.9</td>
<td>41.8</td>
<td>35.7</td>
<td>34.7</td>
<td>34.6</td>
<td>34.6</td>
<td>34.6</td>
<td>34.6</td>
</tr>
<tr>
<td>45</td>
<td>51.8</td>
<td>37.4</td>
<td>30.0</td>
<td>28.2</td>
<td>27.9</td>
<td>27.9</td>
<td>27.9</td>
<td>27.9</td>
</tr>
<tr>
<td>60</td>
<td>50.4</td>
<td>35.5</td>
<td>27.6</td>
<td>25.3</td>
<td>23.8</td>
<td>23.1</td>
<td>21.2</td>
<td>21.2</td>
</tr>
<tr>
<td>75</td>
<td>50.0</td>
<td>34.8</td>
<td>26.3</td>
<td>22.9</td>
<td>20.5</td>
<td>19.4</td>
<td>16.5</td>
<td>16.5</td>
</tr>
<tr>
<td>90</td>
<td>49.0</td>
<td>34.4</td>
<td>25.4</td>
<td>21.1</td>
<td>18.0</td>
<td>16.6</td>
<td>12.8</td>
<td>12.8</td>
</tr>
</tbody>
</table>

Table 4-1: Data reduction and accuracy in building (from 37,875 points)

<table>
<thead>
<tr>
<th>T1</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>69.5</td>
<td>65.5</td>
<td>65.5</td>
<td>65.5</td>
<td>65.5</td>
<td>65.5</td>
<td>65.5</td>
<td>65.5</td>
</tr>
<tr>
<td>10</td>
<td>44.3</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
</tr>
<tr>
<td>15</td>
<td>40.8</td>
<td>19.3</td>
<td>19.1</td>
<td>19.1</td>
<td>19.1</td>
<td>19.1</td>
<td>19.1</td>
<td>19.1</td>
</tr>
<tr>
<td>30</td>
<td>38.4</td>
<td>13.3</td>
<td>12.8</td>
<td>12.7</td>
<td>12.7</td>
<td>12.7</td>
<td>12.7</td>
<td>12.7</td>
</tr>
<tr>
<td>45</td>
<td>37.0</td>
<td>10.5</td>
<td>9.6</td>
<td>9.4</td>
<td>9.3</td>
<td>9.3</td>
<td>9.1</td>
<td>9.1</td>
</tr>
<tr>
<td>60</td>
<td>36.6</td>
<td>9.2</td>
<td>8.0</td>
<td>7.4</td>
<td>7.2</td>
<td>7.2</td>
<td>6.7</td>
<td>6.7</td>
</tr>
<tr>
<td>75</td>
<td>36.5</td>
<td>8.6</td>
<td>7.0</td>
<td>6.0</td>
<td>5.7</td>
<td>5.7</td>
<td>4.9</td>
<td>4.9</td>
</tr>
<tr>
<td>90</td>
<td>36.4</td>
<td>8.1</td>
<td>6.2</td>
<td>5.0</td>
<td>4.5</td>
<td>4.5</td>
<td>3.6</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Table 4-2: Data reduction and accuracy in terrain (from 45,126 points)

Figure 4-11: Results by applying different height differences (example)
However, from user’s point of view, it will be more understandable if the results are provided on the basis of remained rate and $RMSE$ together with two threshold values. That is because a user may be more interested in how the accuracy will change after applying the reduction method. Table 4-3 classifies the results into 11 classes based on the remained rate. The combinations of two thresholds are selected from which have the smallest $RMSE$ within similar remained rate. All the combinations of two threshold values shown in Table 4-3 are selected in this way, and all charts used when selecting threshold values can be referred to appendix A.

From Table 4-3, a user can estimate how much the data can be reduced within the ranges of desired $RMSE$. For example, the data should not be reduced to less than 20% in order to get the data set which has less than 100 cm differences in the building and 7 cm difference in the terrain. If the data is reduced to less than 20%, then the height differences will become more than 100 cm or 7 cm respectively. In addition, a user can adjust one of threshold values in each class to get more sensitive points by that specific parameter, even though he can not estimate exactly how the results will be affected. He only
can estimate roughly that the data will be more reduced and \( RMSE \) will be increased by increasing one of threshold values.

In general, the smaller \( RMSE \) can be obtained with more remained points since they show in proportional relationship as was seen in the most cases in Table 4-1 and Table 4-2. However, some opposite cases are also detected in the several combinations of threshold values in terrain data. The smaller \( RMSE \) was obtained with smaller numbers of points: i.e. \( RMSE \) (0.05) in remained rate 29% with \( T1=1 \) and \( T2=10 \) is smaller than \( RMSE \) (0.07) with remained rate 44% with \( T1=0.5 \) and \( T2=10 \).

In the following Figure 4-13, let’s say that the dotted lines are linear interpolated surfaces with 3 points in Figure 4-13 (a) and with 5 points in Figure 4-13 (b) respectively, then the difference between this linear surfaces (dotted line) and real surfaces (black line) can be interpreted as \( RMSE \), showing bigger in Figure 4-13 (b) rather than Figure 4-13 (a).

![Figure 4-13: Surface representations](image)

Concerning the impacts of different types of objects, the data could be reduced more accurately in the terrain rather than building by proposed method. This can be clearly recognized by the comparison of \( RMSE \) within the same class. However, this does not mean that points on the building can not be reduced accurately. Instead, it can be explained as the method of accuracy assessment by subtraction of two DEMs may not be an appropriate method for the building which has big height jumps. This matter will be dealt with in chapter 5 again.

### 4.3.3. Visual check

The threshold values in the previous section are selected on the basis of the targeted remained rate and \( RMSE \). However, it is still not clear that which properties of objects are remained in each class, even though the remained rate and \( RMSE \) can be recognized. So, in this section, the reduced point set by two threshold values from Table 4-3 are visualized by using PCM software to analyze the properties of objects in each class from Figure 4-13.
In the case of building, the outer boundaries of building and intersections can be clearly distinguishable in the class 2, and more points from the edges and intersections can be remained by increasing the remained points. The ridges can be distinguishable in the class 6 with more than 25% of original points. From the class 6 and above, points on the roof surfaces are added to the reduced points. The result of visualization of these changes can be referred to Table 4-4, Figure 4-13, and Appendix B for more details.

<table>
<thead>
<tr>
<th>Class</th>
<th>Remained rate (%)</th>
<th>RMSE (m)</th>
<th>T1 (m)</th>
<th>T2 (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer boundaries</td>
<td>5 ~ 10</td>
<td>2.39</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>All edges / ridges</td>
<td>25 ~ 30</td>
<td>0.42</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>Sloped surfaces</td>
<td>35 ~ 40</td>
<td>0.30</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Planar surfaces</td>
<td>50~</td>
<td>0.07</td>
<td>0.5</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 4-4: Classification table for Building

In the case of terrain, more accurate DEM can be generated with relatively less data as was explained before. The boundaries of the objects and the intersections between objects and grounds can be distinguished even in the class 1; points on the sloped area are remained from the class 2; and points from the planes start to remain in the class 4. Table 4-5 and Figure 4-14 provide the classification of these remained surface properties after visual checking, and all visualizations can be referred to Appendix B.

<table>
<thead>
<tr>
<th>Class</th>
<th>Remained rate (%)</th>
<th>RMSE (m)</th>
<th>T1 (m)</th>
<th>T2 (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All edges</td>
<td>0 ~ 5</td>
<td>0.5</td>
<td>7</td>
<td>75</td>
</tr>
<tr>
<td>Sloped surfaces</td>
<td>5 ~ 10</td>
<td>0.18</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>Planar surfaces</td>
<td>15 ~ 20</td>
<td>0.07</td>
<td>1</td>
<td>45</td>
</tr>
</tbody>
</table>

Table 4-5: Classification table for terrain
### Figure 4-14: Visualization of classification table

<table>
<thead>
<tr>
<th>Class</th>
<th>Building</th>
<th>Terrain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer boundaries</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>All edges</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>Sloped surfaces</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>Planar surfaces</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
</tbody>
</table>
5. Discussion

This chapter presents discussions on the findings and limitations faced during the research.

5.1. On neighbourhood problem

The neighboured points used in calculating the value of certain point were 5 points which are selected on the basis of 2 dimensions, especially on the horizontal plane. The reason that 2 dimensions were adopted is to include the points on the roof and the ground points together as its neighboured points. Only the 2 dimensional distances is considered to be selected as neighboured points. Actually, 2 dimensions or 3 dimensions do not make big differences in continuous data set like terrain, but they make big differences in discrete object like building as can be seen in Figure 5-1. So, it can be concluded as 2 dimensional neighbourhood is more appropriate in the data set which is composed of lots of objects with big height jumps like building.

However, the real point data by airborne laser scanner are not distributed equally, and even gaps exist by various reasons like existence of non-reflecting surfaces or occlusion by objects. In the case of gaps, most neighboured points around gaps may be selected from only one side of that seed point as can be seen in Figure 5-2. In this case, the calculated values by proposed parameters will not correctly reflect the properties of surfaces for this kind of data set. The k-nearest neighbourhood with angle criterion suggested by (Linsen, 2001) can be considered as one of the solutions for this problem.
Another important issue when defining the neighbourhood is the number of neighboured points. 5 neighboured points were used for the calculation of values in this research since the calculated values did not show big differences according to the number of neighboured points in the given data set. However, this is not always true. The number of neighboured points can influence the resulting values a lot in some cases. For example, the small number of neighboured points can reflect the properties of local surrounding more sensitively since the value of point will be averaged by the number of neighbours. However, this can be also opposite. In the area around gaps again, the selected points by large number of neighboured points can include the points from the other side of the gaps which may have very big different properties. In this case, the calculated values by large numbers of neighboured points will be much higher than the value by smaller number of neighboured points. So, we can say that the selection of numbers of neighbours is dependent on the area where the points were taken. In the data set with lots of gaps, the large number of neighboured points would be recommended, while in opposite case small numbers of neighboured points would be preferred.

5.2. Difference of DEM as a measure of accuracy assessment for building?

For the accuracy assessment, DEM was used in both for the terrain points and building points. DEM is normally used as a way of representing the height information of continuous fields like terrains. However, in the area with discrete objects like buildings or man-made creatures, DEM can not be generated accurately. This can be clearly noticed from the results in chapter 4. The calculated RMSE of DEMs subtraction in building shows quite larger ranges (from 0.07 to 2.39) compared to the ranges in terrain (from 0.01 to 0.95). Especially, the large RMSE appears along the edges of roof as can be seen in Figure 5-3.

![Figure 5-3: Visualization of differences of DEM subtractions (example)](image)

Actually, several interpolation methods can be adopted when generating DEM out of TIN. The adopted interpolation method in this research was Natural neighbour interpolation. However, the cells on the edges can never have accurate values by any methods since some part of a cell on the edge will always lie on the roof and some part will lie on the grounds in any cases, being represented by one value. This is not a matter of selection of interpolation methods, but an inherent characteristic in DEM generation as a raster format. So, RMSE by DEM subtraction in urban area may not be a good
indicator for the accuracy assessment. However, the values on the roof surface or planar parts are still reliable even in such areas, showing white colour.

5.3. On the threshold values

The threshold values used in the test were selected on the basis of the calculated values by two parameters. By simply increasing the threshold values, the more significant points can be remained. However, the concepts of significance can not be applied in a same way to all applications. Significant points in one application may not be regarded as significant points in a same way by other applications. For example, the points on the planar surfaces may be redundancies in 3D modelling, but may be significant points in visualization of surface roughness. So, the most important thing in the data reduction may be how to give the threshold values on the calculated values. In the following, two ways of determining threshold values are discussed.

One way of selecting threshold values is to give the ranges of threshold values from the visualized values. This is exactly how the first proposed method in this research used. In the proposed method, the calculated values are showing from red to cyan which matches from 180 to 0 degree. A user can specify the threshold values by selecting colours scheme in terms of angles. In the proposed method, the ranges are only selected from which is ‘greater than some values’. However, the ranges are not necessarily to be like this, instead they can be selected in any part of them by using appropriate combinations of relational inequality together with logical connectives with two parameters based on the applications.

Another way to select the threshold values is to utilize the classified results in Table 4-3, Table 4-4, and Table 4-5. Even though the tests have been done only with one building and one terrain data, the threshold values in each class can be used as standards when defining the threshold values. The calculated values may show small differences according to the data properties. However, if the similar ranges of threshold values are applied, at least similar types of features will be selected regardless of the reduction rate or RMSE.
6. Conclusions and recommendations

This research aimed at developing a new method which can remove non-significant points out of dense point clouds according to the significance values calculated by two defined parameters. Accordingly, the main focus has been put on deriving and formulating the criteria to calculate the significance values of points. Two parameters were derived for calculating significance values of the points in this research, and were assessed by analyzing the results of two DEMs subtraction. In this chapter, the answers for the proposed research questions will be presented and some recommendations for further research will be proposed.

6.1. Conclusions

Five research questions were proposed in the chapter 1.

The first and second questions were ‘How can parameters be defined in terms of surfaces representation? / How can the significance value of each point be determined based on the pre-defined parameters?’

To answer these two questions, two parameters are defined to represent the surface properties, which are sum of height differences with the constraint of maximum slope 3 (70°) and the change of normal. Both values are calculated within 2 dimensional neighboured points separately, and then combined for selection. In particular, normal vectors of a point with respect to tangent plane at that point were calculated by performing the eigenvector analysis.

The third and fourth research questions were ‘What are the impacts of different types of surfaces and different point density on the results?’

Basically, two point data sets, building and terrain with similar point density were put into test in this research, showing better results with the terrain data rather than the building. However, this results can be said only when the comparison of two DEMs are used as a method for accuracy assessment. If the other method is applied for accuracy assessment, the opposite results can be derived also. In addition, point density only affects the selected numbers of neighboured points and thus may result in different values. So, the way of defining the neighboured points may be more influencing factor rather than the surface properties or point density.

The fifth research question was whether the method can remove all points except for building corner points. Theoretically, only the corner points located on the outer boundaries of the building may have highest values than any other points in the building except for the points on the peak. That is because the neighboured points of outer corner point are composed of 3-quarters from grounds and 1-quarter from the roof, while the neighboured points for the edges points are composed of a half from grounds
and a half from roof. However, the corner points located at the ends of ridges can not be differentiated by the proposed method.

Basically, the proposed method is designed to select points with higher values than defined threshold values. However, the method is quite flexible for users to select the points interactively by adjusting the threshold values on their interests. In addition, the comparison of DEM subtraction as measure of accuracy assessment can be a quite useful method at least for the terrain data when making DTM

### 6.2. Recommendations

The proposed method removes the points according to the calculated values. So, the most important thing to improve the quality of reduced point set is to get rid of factors which can give bad influences on the results. In that sense, if the neighbourhood can be defined in a proper way when encountering to gaps, the results can be much more improved.

Regarding accuracy assessment of the reduced points, RMSE of DEMs subtractions may not be a good indicator as a measure of accuracy assessment in case of building which is composed of points with big height jumps. RMSE appears large along the edges of the building as was explained in previous chapter. So, the further research should be on the development of the appropriate assessment method for the objects like building.

In addition, the threshold values were selected based on the analysis of the calculated values with one building data and one terrain data. However, to make the threshold values more applicable to other data set, the more tests with more diverse data set should be done.
Reference


BRODSKY, D. & WATSON, B. (2000) 'Model Simplification Through Refinement'. GRAPHICS INTERFACE '00, pp.221-228


Appendix A: Remained rate and RMSE

Building (different sum of height difference with fixed change of normal)
Building (different change of normal with fixed height difference)
Terrain (different sum of height difference with fixed change of normal)
Terrain (different change of normal with fixed height difference)

DATA REDUCTION OF POINT CLOUDS ACQUIRED BY AIRBORNE LASER SCANNING
## Appendix B : Visualization of remaining points

<table>
<thead>
<tr>
<th>Class</th>
<th>Visualization (building)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Oblique view</td>
<td>Top view</td>
</tr>
</tbody>
</table>
| 2     | ![Image](image1.png)    | ![Image](image2.png) | - Building outer boundary  
- Intersection between building and grounds |
| 3     | ![Image](image3.png)    | ![Image](image4.png) | - Building outer boundary (more)  
- Intersection between building and grounds (more)  
- little points from the roof |
| 4     | ![Image](image5.png)    | ![Image](image6.png) | - Building outer boundary  
- Intersection between building and grounds  
- little points from the roof |
| 5     | ![Image](image7.png)    | ![Image](image8.png) | - Building outer boundary  
- Intersection between building and grounds  
- little points from the roof |
| 6     | ![Image](image9.png)    | ![Image](image10.png) | - Building outer boundary  
- Intersection between building and grounds  
- little points from the roof  
- Ridge of building |
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 7 | ![Image](image1.png) | - Building outer boundary  
- intersection between building and grounds  
- little points from the roof  
- Ridge of building  |
| 8 | ![Image](image2.png) | - Building outer boundary  
- intersection between building and grounds  
- Ridge of building  
- **Some points from the roof surfaces**  |
| 9 | ![Image](image3.png) | - Building outer boundary  
- intersection between building and grounds  
- Ridge of building  
- **More points from the roof surface**  |
| 10 | ![Image](image4.png) | - Building outer boundary  
- intersection between building and grounds  
- Ridge of building  
- More points from the roof surface  |
| 11 | ![Image](image5.png) | - Building outer boundary  
- intersection between building and grounds  
- Ridge of building  
- More points from the roof surface  |
| 12 | ![Image](image1.png) | - Building outer boundary  
- intersection between building and grounds  
- Ridge of building  
- More points from the roof surface |
| 13 | ![Image](image2.png) | - Building outer boundary  
- intersection between building and grounds  
- Ridge of building  
- More points from the roof surface  
- Some points from the ground |
<p>| 0 | <img src="image3.png" alt="Image" /> | Original data |</p>
<table>
<thead>
<tr>
<th>Class</th>
<th>Visualization (terrain)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Oblique view</td>
<td>Top view</td>
</tr>
</tbody>
</table>
| 1     | ![Image] | ![Image] | - Boundary of objects  
- Intersection between object and grounds |
| 2     | ![Image] | ![Image] | - Boundary of objects  
- Intersection between object and grounds  
- **Some points from the sloped area** |
| 3     | ![Image] | ![Image] | - Boundary of objects  
- Intersection between object and grounds  
- More points from the sloped area |
| 4     | ![Image] | ![Image] | - Boundary of objects  
- Intersection between object and grounds  
- More points from the sloped area  
- **Some points from the planes & roads** |
| 5     | ![Image] | ![Image] | - Boundary of objects  
- Intersection between object and grounds  
- More points from the sloped area  
- More points from the planes & roads |
| 6  | ![Image 6](image1) | - Boundary of objects  
- Intersection between object and grounds  
- More points from the sloped area  
- More points from the planes & roads |
| 8  | ![Image 8](image2) | | ![Image 8](image3) |
| 9  | ![Image 9](image4) | - Boundary of objects  
- Intersection between object and grounds  
- More points from the sloped area  
- More points from the planes & roads |
| 13 | ![Image 13](image5) | - Boundary of objects  
- Intersection between object and grounds  
- More points from the sloped area  
- More points from the planes & roads |
| 0  | ![Image 0](image6) | Original data |