

# **Radiometric correction of mobile laser scanning intensity data**

Donghyun Oh  
February, 2010

# Radiometric correction of mobile laser scanning intensity data

by

*Donghyun Oh*

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. M.G. Vosselman

External examiner: Dr. B. Höfle

Supervisor: Dr. M. Rutzinger

Second supervisor: Prof. Dr. M.G. Vosselman



**INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION  
ENSCHEDA, THE NETHERLANDS**

### **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 recent development of laser scanning technique has made it possible to record the intensity of received signals as well as to provide the accurate geometric information of a point cloud. Many studies show the potential of intensity data for a lot of applications (e.g. strip adjustment, segmentation and feature extraction). Meanwhile, the problems in intensity-based applications have been presented due to radiometric systematic bias and reflectance noise. Traditionally, the intensity has been corrected by the radar range equation in airborne laser scanning (ALS) data. However, it is found that these models are not sufficient to correct terrestrial laser scanning (TLS) data. Moreover, mobile laser scanning (MLS) data have not been thoroughly studied.

This research proposes the correction model of MLS intensity data to be able to reduce its radiometric systematic bias by two main approaches: a) theoretical model-based approach and b) empirical model-based approach. The empirical model-based approach is tested by two models: adjusted radar range equation and data fitting by polynomials. In the initial phase filtering of multi-echoes and outliers is carried out which are not related to radiometric systematic noise. The geometry of a remaining point cloud is reconstructed into the range and incidence angle which are the main influencing factors on intensity data. In the following phase the theoretical model-based approach corrects intensities by range and incidence angle that individual points have. The empirical model-based approach uses the correction functions dependent on the range and incidence angle, whereby the parameters are defined by samples from homogeneous surfaces.

The evaluation is performed based on (a) consistency of intensities in a homogeneous surface, (b) consistency in the two sensors, (c) contrast between different surfaces and (d) local noise reduction of intensities. In the assessment result it is indicated that the radar range equation is not suitable for MLS data. Furthermore, it is enlightened that among tested methods the data fitting by polynomials clearly reduces the intensity variation in a homogeneous surface and different sensors, and makes the contrast significant between different surfaces. Also shown is that the correction by data fitting can be used in the applications such as feature extraction, classification and segmentation.

**Keywords:** mobile laser scanning, intensity, reflectance, radiometric correction, calibration



## Acknowledgements

---

I would like to give my sincere thanks to two supervisors. Dr. M. Rutzinger, my first supervisor, had a close look to my research and supported me with advices apt for the occasion. This thesis would not have been possible to go in right directions without his constructive and tireless guidance. Prof. Dr. M.G. Vosselman, my second supervisor, made me feel interest on laser scanning field and provided detailed insight. His clear instruction will be kept as the symbol of ITC in my mind. Especially, I appreciate for two supervisors to always give me the time for discussion despite their busy schedules whenever I asked meeting.

I owe my deepest gratitude to TopScan GmbH to provide the laser scanning data, and Dr. B. Höfle to organize the experiment by reference target that I really wished to do.

I want to extend my thanks to all of classmates to offer their board experience and discuss with me during 18 months. They encouraged me to be able to overcome my linguistic problem and narrow background on this field. I would like to pay special thanks to our class representative, Mr. Malakai Vakautawale, and research theme (TOPMAP) members, Arun, Janak and Yuxian.

I would like to show my gratitude to Republic of Korea Army (ROKA) to serve full support and thoughtful consideration to enable me to study in Geo-Information Science and Earth Observation (ITC) and live in the Netherlands. I am deeply indebted to ROKA and will try to do my best to return their favour with the wider knowledge.

Last but not least, I would like to give my best thanks to my wife, Misun Park, whose unlimited love and warm devotion enabled me to cope with one of challenges in my life. Since she has shown only delightful faces despite a difficult foreign life, I could concentrate on this research and complete the thesis successfully. Also, I want to express my thanks to my daughter Yunim to grow healthily in the Netherlands without any problem.

석사과정을 위해 아낌없는 배려와 관심을 보내주신 모든 분들께 감사드립니다.



# Table of contents

---

<b>1. Introduction</b> .....	<b>1</b>
<b>1.1. Motivation and problem statement</b> .....	<b>1</b>
<b>1.2. Research identification</b> .....	<b>2</b>
1.2.1. Research objectives .....	3
1.2.2. Research questions .....	3
1.2.3. Innovation aimed at .....	4
<b>1.3. Methodology</b> .....	<b>4</b>
<b>1.4. Structure of thesis</b> .....	<b>5</b>
<b>2. Related techniques for correction of intensity data</b> .....	<b>7</b>
<b>2.1. Introduction</b> .....	<b>7</b>
<b>2.2. Correction methodologies of intensity data</b> .....	<b>8</b>
2.2.1. Method based on the theoretical model .....	8
2.2.2. Method based on the empirical model .....	10
<b>2.3. Experiments by reference targets</b> .....	<b>13</b>
2.3.1. Influence by incidence angle (Kukko et al., 2008) .....	14
2.3.2. Influence by sensor own property based on range (Pfeifer et al., 2008) .....	15
<b>2.4. Summary</b> .....	<b>16</b>
<b>3. Proposed Method</b> .....	<b>17</b>
<b>3.1. Overview</b> .....	<b>17</b>
<b>3.2. Pre-processing</b> .....	<b>18</b>
<b>3.3. Reconstruction of scan geometry</b> .....	<b>18</b>
3.3.1. Range .....	19
3.3.2. Surface normal and Incidence angle .....	20
<b>3.4. Data correction based on radar range equation</b> .....	<b>21</b>
3.4.1. Definition of parameters .....	22
3.4.2. Simulation by simplified radar range equation .....	22
<b>3.5. Data correction based on empirical models</b> .....	<b>23</b>
3.5.1. Analysis of uncorrected intensities .....	23
3.5.2. Adjustment of the radar range equation .....	25
3.5.3. Data fitting .....	26
<b>3.6. Summary</b> .....	<b>27</b>

<b>4. Implementation and results.....</b>	<b>29</b>
<b>4.1. Description of dataset.....</b>	<b>29</b>
4.1.1. Dataset and study area.....	29
4.1.2. Examination of test data.....	31
<b>4.2. Pre-processing.....</b>	<b>32</b>
4.2.1. Implementation.....	32
4.2.2. Results.....	32
<b>4.3. Reconstruction of scan geometry.....</b>	<b>33</b>
4.3.1. Sensor trajectory.....	33
4.3.2. Surface normal.....	34
4.3.3. Range.....	35
4.3.4. Incidence angle.....	35
<b>4.4. Data correction based on radar range equation.....</b>	<b>36</b>
<b>4.5. Data correction based on empirical models.....</b>	<b>37</b>
4.5.1. Analysis of uncorrected intensities.....	37
4.5.2. Adjusted radar range equation.....	40
4.5.3. Data fitting.....	43
<b>4.6. Concluding remarks.....</b>	<b>46</b>
<b>5. Evaluation and analysis.....</b>	<b>47</b>
<b>5.1. Statistical analysis based on samples.....</b>	<b>47</b>
5.1.1. Evaluation scheme.....	47
5.1.2. Results analysis.....	48
<b>5.2. Analysis by classification of surface types.....</b>	<b>54</b>
5.2.1. Evaluation scheme.....	54
5.2.2. Results analysis.....	54
<b>5.3. Concluding remarks.....</b>	<b>56</b>
<b>6. Conclusions and recommendations.....</b>	<b>57</b>
<b>6.1. Conclusions.....</b>	<b>57</b>
<b>6.2. Recommendations.....</b>	<b>58</b>
<b>6.3. Further study.....</b>	<b>59</b>
<b>7. Reference.....</b>	<b>61</b>
<b>Appendix A: Correction of test data 2 by the radar range equation.....</b>	<b>63</b>

## List of figures

---

<i>Figure 1-1: Laser scanner geometry (Ohio, 2008)</i> .....	2
<i>Figure 1-2: Applied workflow</i> .....	4
<i>Figure 2-1: Measured (left) and corrected (right) result (Coren and Sterzai, 2006)</i> .....	10
<i>Figure 2-2: Recorded and empirically corrected intensities (Höfle and Pfeifer, 2007)</i> .....	11
<i>Figure 2-3: Intensities for incidence angle (Jutzi and Gross, 2009)</i> .....	13
<i>Figure 2-4: The effect of incidence angle for surface roughness (Kukko et al., 2008)</i> .....	14
<i>Figure 2-5: The effect of incidence angle for surface brightness (Kukko et al., 2008)</i> .....	14
<i>Figure 2-6: Mean intensities for Riegl and Optech at different distances (Pfeifer et al., 2008)</i> .....	15
<i>Figure 3-1: Workflow of proposed method</i> .....	17
<i>Figure 3-2: Workflow for reconstruction of scan geometry</i> .....	18
<i>Figure 3-3: Description of range calculation (points were represented by intensity value.)</i> .....	19
<i>Figure 3-4: Definition of incidence angle (profile view - represented by intensity value)</i> .....	20
<i>Figure 3-5: Simulation of intensity data by radar range equation (R: range, A: incidence angle)</i> .....	22
<i>Figure 3-6: Description of used terms in sampling procedure</i> .....	23
<i>Figure 4-1: Offset of LYNX Mobile Mapper (Background image: (Optech, 2007))</i> .....	29
<i>Figure 4-2: Test data 1 – road corridor with several planer surfaces with homogeneous brightness</i> .....	30
<i>Figure 4-3: Test data 2 – Paved road surface with points of various ranges and incidence angles</i> .....	31
<i>Figure 4-4: Road and road marks classified by different thresholds (a) 250~1000 and (b) 150~100031</i>	
<i>Figure 4-5: Histogram by intensities of test data 1 (red line: threshold, 95%)</i> .....	32
<i>Figure 4-6: Filtering results of test data 1 (color-coded by filtering scheme)</i> .....	32
<i>Figure 4-7: Estimated sensor trajectory (a) without filtering stopping points and (b) after filtering</i> .....	33
<i>Figure 4-8: Estimated surface normal (color-coded by the angle with vertical axis)</i> .....	34
<i>Figure 4-9: Estimated ranges (color-coded by ranges, blue line: sensor trajectory)</i> .....	35
<i>Figure 4-10: Estimated incidence angles (color-coded by incidence angles)</i> .....	35
<i>Figure 4-11: Correction results by theoretical model</i> .....	36
<i>Figure 4-12: Sampling for defining empirical model (test data 1)</i> .....	37
<i>Figure 4-13: Sampling for defining empirical model (test data 2)</i> .....	38
<i>Figure 4-14: Correction results by adjusted radar range equation</i> .....	42
<i>Figure 4-15: Scatter plot by range and processed relative intensity (red *: asphalt, green+: road marks, blue ★: pedestrian road, black ◇: building wall and magenta △: building roof)</i> .....	43

*Figure 4-16: Fitting lines on both data*..... 44

*Figure 4-17: Correction result by data fitting* ..... 45

*Figure 5-1: Sampling for evaluation*..... 49

*Figure 5-2: Comparison of sample groups in homogeneous area* ..... 51

*Figure 5-3: Standard deviations within sample groups*..... 53

*Figure 5-4: Classification result of road marks* ..... 55

*Figure 5-5: Example of a classified detailed object (a) image by digital camera, (b) uncorrected data and (c) corrected data*..... 55

*Figure 6-1: MLS intensity experiment by reference target (Höfle, 2010)* ..... 59

*Figure 6-2: Scatter plots of experiments (red\*: strip 1 of sensor 1, magenta○: strip 2 of sensor 1, blue+: strip 1 of sensor 2, black△: strip 2 of sensor 2)*..... 60

## List of tables

---

<i>Table 2-1: Relative intensity value in ALS and TLS in low angles of incidence (Kukko et al., 2008) ..</i>	<i>15</i>
<i>Table 3-1: Equations and surface normal adopted according to the arrangement of neighborhood...</i>	<i>21</i>
<i>Table 3-2: A table of atmospheric losses for the wavelength of 1,550 nm (Kim et al., 2001).....</i>	<i>22</i>
<i>Table 3-3: Generic example of lookup table of intensity with respect to signal power .....</i>	<i>24</i>
<i>Table 4-1: The number and proportion of filtered points in test data 1 .....</i>	<i>33</i>
<i>Table 4-2: Box plots by sample sets (data set 1).....</i>	<i>39</i>
<i>Table 4-3: Box plots by sample sets (data set 2).....</i>	<i>40</i>
<i>Table 4-4: Estimated parameters according to sample sets .....</i>	<i>41</i>
<i>Table 4-5: Applied equations according to range .....</i>	<i>41</i>
<i>Table 4-6: Correction model based on data fitting .....</i>	<i>44</i>
<i>Table 5-1: Distribution of medians of ranges and incidence angles.....</i>	<i>49</i>
<i>Table 5-2: Distribution of intensities (median and standard deviation).....</i>	<i>52</i>



# 1. Introduction

## 1.1. Motivation and problem statement

With the development of sensor techniques such as laser scanning, wider areas with high point densities can be surveyed. Beside the geometric information (x, y, z) the provision of intensity data ( $I$ ) has extended laser scanning applications. Recently, it has been used in a wide range of applications such as forestry, glacial research as well as 3D city modelling. The intensity has been studied as two main categories a) sensor related procedures such as estimation of adjustment between laser scanner strips and b) surface segmentation and classification methods (Höfle and Pfeifer, 2007). Moreover, it was found that the ranging error is caused from different signal-to-noise ratios depending on signal powers. It is proposed that the intensity based application can be used in geometric calibration procedures since the travel-time of emitted pulses is determined by returned echo amplitude (Pfeifer et al., 2008).

The intensity, physically the power of incoming echoes, has been widely used as synonym for the amplitude, reflectance or energy in the terminology of laser scanning (Jutzi and Gross, 2009). However, those terms are not correct expressions for the intensity used in this study since most laser scanners record the signal power at a certain moment as a relative value of 8 or 16 bit. Also, the intensity data cannot be applied in further applications without correction since it has the systematic bias depending on ranges and incidence angles. This is demonstrated by surfaces with the same property which are represented as different values in the uncorrected intensity data. Its causes were

theoretically described by  $P_r = \frac{P_t D_r^2}{4\pi R^4 \beta_t^2} \eta_{sys} \eta_{atm} \sigma$ , radar range equation (Jelalian, 1992) where

$P_r$  (received signal power) is a function of  $P_t$  (transmitted signal power),  $D_r$  (receiver aperture diameter),  $R$  (range from sensor to target),  $\beta_t$  (laser beam width),  $\sigma$  (target cross-section),  $\eta_{sys}$  and  $\eta_{atm}$  (system and atmospheric transmission factor). Under the assumption that a dataset is collected from one sensor, transmitted pulses are constant, footprints on reflected objects are circular, surface has Lambertian characteristics and atmospheric condition from sensor to target is constant, the

equation can be simplified into  $P_r \propto \frac{\rho}{R^2} 10^{-2Ra/10000} \cdot \cos \alpha$ , which is a proportional function of  $R$ ,

$\alpha$  (incidence angle),  $a$  (air attenuation) and  $\rho$  (target reflectance). Since  $R$  and  $\alpha$  can be calculated from laser scanner geometry (Figure 1-1) and  $a$  can be taken from physical experiment data of laser beam propagation in a certain weather condition (Jelalian, 1992; Kim et al., 2001), relative radiometric correction algorithms were developed using this simplified radar range equation. Furthermore, empirical models derived from the data itself (e.g. fitting polynomial and logarithmic equation and Taylor series expansion) were used to define the relation between the influencing factors and the received intensity since the assumptions in the procedure to simplify radar range equation are not kept in all point cloud data (e.g. elongated pulses tilted on targets, atmospheric condition from sensor to target not to be constant and different sensor types).

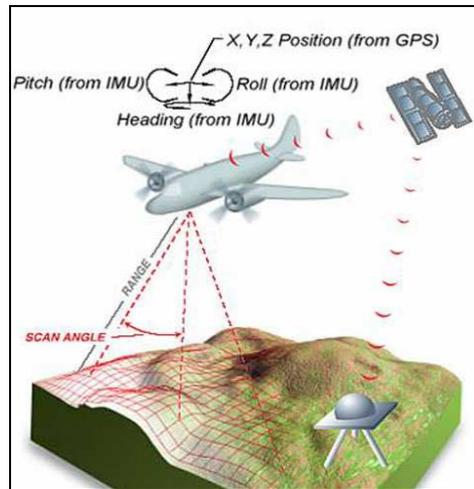


Figure 1-1: Laser scanner geometry (Ohio, 2008)

In the year 2000 one of vehicle-borne laser mapping systems was developed as one of the experiments in the university of Tokyo (Manandhar and Shibasaki, 2000). Recently, dynamic MLS (mobile laser scanner), a synonym to vehicle-borne terrestrial laser scanner, has been developed which uses integrated GPS/ IMU units and one or more sensors with high frequency. Most MLSs carry also multiple video or small-format digital cameras. MLS is becoming an operationally used technique and is being used for surveying and mapping of road corridors and urban areas. Recent MLS systems have a data collection rate of more than 1 million points per second (Shan and Toth, 2008). They are continuously developed. For example, MLS compared to ALS (airborne laser scanner) collects point clouds from relatively short ranges and various incidence angles. The LYNX system, one of MLS sensors, takes data within the restricted range of 100m, FOV of 360° and point density of 1 point per centi-meter at 100 km/h (Optech, 2007).

Some research has been done on intensity correction so far. However, the current correction algorithms of intensity data have been experimented with TLS (terrestrial laser scanner) data (Pfeifer et al., 2007; Pfeifer et al., 2008), ALS data (Coren and Sterzai, 2006; Höfle and Pfeifer, 2007) or both (Kaasalainen et al., 2009; Kaasalainen et al., 2008). Little research was done on MLS data. The effect of influencing factors on intensity data in MLS will be different with that in other devices. This research is motivated toward developing the correction methods for MLS data, which will benefit for many applications such as segmentation and classification.

## 1.2. Research identification

In this research a radiometric correction model for intensity data is developed by two approaches.

Firstly, intensity will be corrected into relative values rather than absolute values. Due to the advantages of the active sensor system which emits pulses in a specific wavelength in near infrared and the fact that reflected surfaces have a unique reflectivity for a specific wavelength, intensity data can be considered to have much more potential. For example, if absolute reflectivity of surfaces is obtained, surface type will be able to be disclosed by lookup table for reflectivity. In spite of this, it is not possible to measure absolute intensity with current MLS systems taken from the field instead of laboratory testing as all required data of theoretical radar range equation cannot be correctly embodied.

For example, even signal power of emitted pulses is unknown and the environmental factors in data collection cannot be exactly measured for all survey areas. Since no ground reference target was available during data collection, it is known that only relative correction is possible (Höfle et al., 2009). However, even if corrected intensity is a relative value, it is still valuable data as referring to that most image processing methods are based on relative reflectance values by image enhancement such as stretching, histogram equalization and filtering.

Secondly, intensity will be processed with the individual point rather than raster-based methods. According to Höfle and Pfeifer (2007), intensity data can be processed in two ways: a) algorithms that individually process each point and b) image processing techniques based on interpolated intensity due to the preference for straightforward raster-based methods. However, if raster-based technique is used in correction algorithm, the quality of the algorithm will deteriorate resulting in loss of detail. MLS maps a 3D environment which cannot be represented in raster domain. Therefore, correction models will be processed and developed with an individual point rather than raster-based technique.

### **1.2.1. Research objectives**

The main objective of this research is to develop the radiometric correction model for intensity data obtained by MLS. This can be sub-divided into the following specific objectives to achieve the main objective.

- i. To reconstruct scan geometry i.e. range and incidence angle which significantly affect MLS intensity and investigate the relation between scan geometry and intensity.
- ii. To correct intensity by two available methodologies: a) radar range equation with theoretical bases and b) empirical models derived from data itself.
- iii. To identify the usability of models in the applications such as segmentation, classification, and feature extraction (road markings).

### **1.2.2. Research questions**

The following questions have to be answered to achieve research objectives.

- i. How can range and incidence angle be reconstructed in MLS data?
- ii. How are range and incidence angle related to MLS intensity?
- iii. How differently are range and incidence angle related to MLS intensity data with respect to different surface types and sensors (left and right)?
- iv. Can the relations be modelled by radar range equation? How can they be modelled by empirical functions?
- v. How can the performance of the algorithm be evaluated?
- vi. Can the corrected results be used in laser scanning applications such as segmentation, classification, and feature extraction?

### 1.2.3. Innovation aimed at

As mentioned, this research starts from the idea that the model for correction of MLS intensity data will be different with that of ALS and TLS intensity data. Therefore, the intended innovations of this research are to develop the new model applying the characteristics of MLS intensity data which has not been studied in previous research so far. This research will be able to contribute to widen the use of MLS intensity data.

### 1.3. Methodology

This research is conducted in following workflow to achieve research objectives (Figure 1-2).

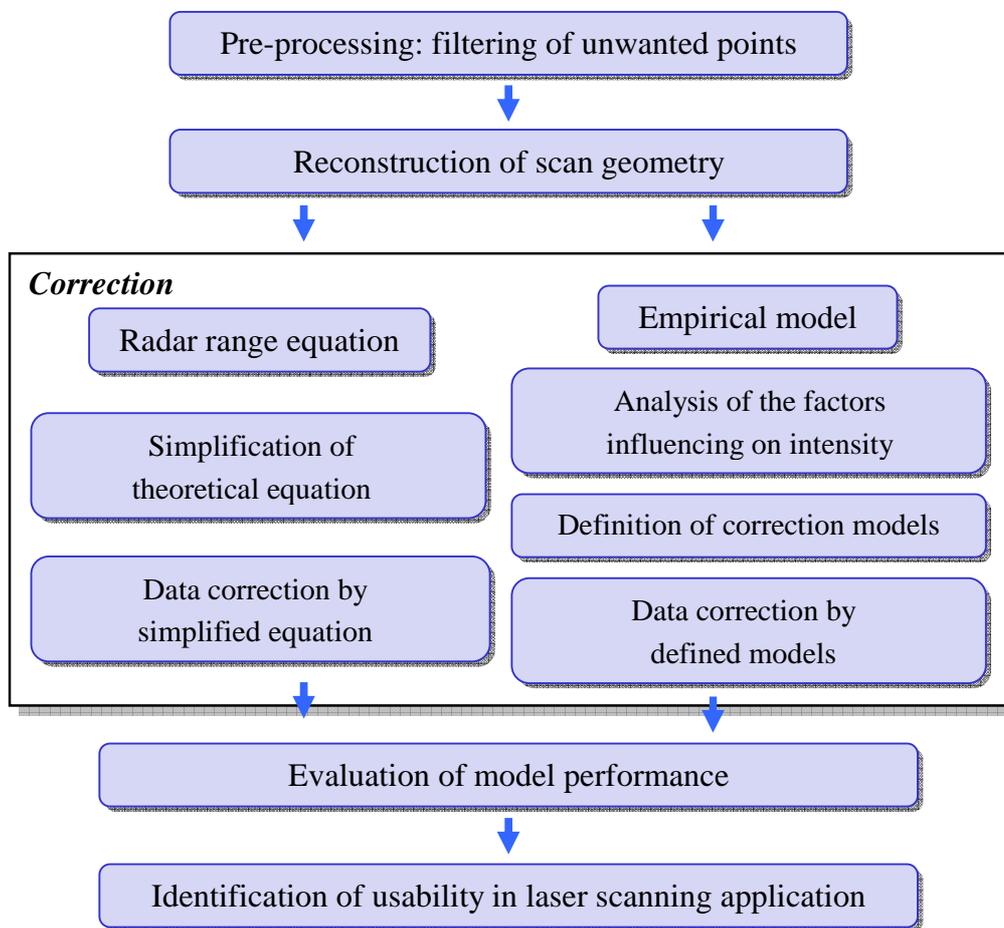


Figure 1-2: Applied workflow

- i. Data preparation: As the point cloud data usually consists of enormous amount of points, unnecessary points for modelling correction algorithm will be filtered out (e.g. multi-echoed pulse and pulse with extreme intensity due to specular reflection or other energy sources).
- ii. Reconstruction of scan geometry: The scan trajectory will be reconstructed from time stamp and 3D coordinate (time, x, y, z) to laser shot vector (range, incidence angle) using laser scanner geometry on point clouds, GPS and offset data of 'GPS centre to sensor projection centre'. The range and incidence angle can be calculated respectively from Euclidean distance between target and sensor and the angle enclosed by surface normal and the vector of 'target

to sensor'. In addition, as the dataset contains information about GPS positions rather than sensor positions, offset data are taken into account in this procedure.

- iii. Correction of intensity data: In this step, intensity data are corrected by two approaches.
  - o Radar range equation: Firstly, correction of intensity data is approached with a well-known theoretical model. However, it makes absolute correction difficult since this equation includes variables that MLS does not provide and the intensity data are relative values rather than absolute values. Therefore, it is simplified to proportional expression by reasonable assumptions. Finally, this simplified equation is used in data correction.
  - o Empirical model: Secondly, the radar range equation was derived from laboratory tests. However, there are lots of additional factors influencing intensity data in reality. Therefore, the empirical models are derived from data itself. Data subsets are sampled from homogeneous surfaces (e.g. road markings, car and bicycle lanes of paved road and building walls), in order to control the effect of physical properties of surfaces in modelling procedure. Based on the inspected effect of influencing factors in samples, correction models are defined by the incorporation of each factor and these models will be used in data correction.
- iv. Evaluation of model performance: As these models are for relative correction, the results corrected by the final models defined in previous step are validated by the comparison of intensities newly sampled from homogeneous area with different geometry after data correction. Due to noise of measurements in homogeneous surfaces of highly dense point clouds, the box plot of each sample is used in the comparison of results (Jutzi and Gross, 2009).
- v. Application of corrected data: Finally, the corrected intensity data are used in classification of surface types such as road marks, pavement types and surface types of building wall.

#### **1.4. Structure of thesis**

This thesis consists of six main chapters. Chapter 1 is a general introduction to the research and describes the main objective, motivation and the proposed research questions. Chapter 2 reviews the state of art in correction algorithms that have been applied to ALS and TLS datasets. Chapter 3 shows the proposed methodology for intensity data correction and chapter 4 presents the study area and the corrected results by proposed methods. The performance and quality of methods are evaluated and discussed in chapter 5. Finally, the thesis is concluded and the recommendations are presented in chapter 6.



## 2. Related techniques for correction of intensity data

### 2.1. Introduction

Up till now, various algorithms have been developed to correct intensity data. The initial attempt was to normalize intensity data with respect to the variation of ranges. In other words, it considers only a range ( $R$ ) as an influencing factor and intensity values ( $I$ ) were normalized to pre-defined standard range ( $R_s$ ) using following normalizing equation (Luzum et al., 2004) and a linear regression model of intensity for a range (Donoghue et al., 2007).

$$I(R_s) = I \cdot \frac{R^2}{R_s^2} \quad (2.1)$$

Since the first correction algorithms were developed for ALS and the influence of the incidence angle is less than in TLS and MLS, it showed improved results. Afterwards, incidence angle, air attenuation as well as a range were considered as influencing factors. Using physical equations, the relative radiometric correction algorithm was proposed to correct the effects of range, incidence angle (exactly scan angle), and air attenuation on intensity (Coren and Sterzai, 2006). Furthermore, the data-driven correction algorithm using the tested empirical model was developed. Under the assumption that returned intensity is proportional to the ground reflectance and related to the flying height via a tested empirical model (inversely quadratic model and Taylor series expansion model), returned intensities were corrected by the statistical model defined with estimated parameters (Höfle and Pfeifer, 2007). The algorithm was developed to add the incidence angle that is calculated by surface normal and the vector from target to sensor (Jutzi and Gross, 2009). Besides, this approach was used to correct MLS intensity in an application for extracting road markings even though it was the simple algorithm by a second order polynomial function to embody only an approximate range derived along the profile (Jaakkola et al., 2008).

Meanwhile, correction algorithms were advanced with more physical approach to investigate the measurements of reference targets in laboratory and field (Kaasalainen et al., 2008; Kukko et al., 2008). Furthermore, as it was disclosed that measured intensities have different patterns in short and long ranges depending on sensor types, the sensor specific correction model was developed. It was suggested that the material properties can be retrieved with only TLS intensity data corrected by the empirical model (Pfeifer et al., 2008).

As mentioned earlier, the correction methods can be classified into two categories (theoretical model-based and empirical model-based methods). The general description and the correction results in ALS dataset by each category are explained in section 2.2. Section 2.3 is devoted for the physical experiments using reference targets to inspect the influence by the range and incidence angle. Finally, the summary and concluding remarks for this research are delivered in section 2.4.

## 2.2. Correction methodologies of intensity data

### 2.2.1. Method based on the theoretical model

This method basically uses the following known general formula, radar range equation (Jelalian, 1992), which defines the relation between the transmitted signal power and received signal power.

$$P_r = \frac{P_t \cdot D_r^2}{4\pi R^4 \beta_t^2} \eta_{sys} \eta_{atm} \sigma \quad (2.2)$$

Where,  $P_r$  and  $P_t$  = received and transmitted signal power

$D_r$  = receiver aperture diameter

$R$  = range from sensor to target

$\beta_t$  = beam width of transmitter

$\eta_{sys}$  and  $\eta_{atm}$  = system and atmospheric transmission factor

$\sigma$  = target cross-section

As shown in equation (2.2), it consists of the three main factors that diminish the transmitted signal power: a) target, b) sensor, and c) atmospheric parameters. The receiver aperture diameter ( $D_r$ ) is size of the hole that can admit laser shots. The beam width ( $\beta_t$ ) is the diameter on perpendicular plane to the beam axis and it is increased with a certain angle (beam divergence). Thus, in this equation it needs to be assumed that the receiver field of view matches the beam divergence and that beam speed is fast enough that the positions of emitter and detector coincide at single signal travelling. However, laser scanning intensity data are collected in dynamic environment rather than laboratory conditions. Also, data themselves are of relative values produced by a sensor specific lookup table. Therefore, it is restricted to correct intensity data into absolute values by using full radar range equation. By the following computational procedure, this equation is simplified to correct intensity data relatively.

The target cross-section ( $\sigma$ ) stands for the target-related factor including all target properties and is defined as the following equation.

$$\sigma = \frac{4\pi}{\Omega} \rho \cdot A_s \quad (2.3)$$

Where,  $\Omega$  = scattering solid angle

$\rho$  = target reflectance

$A_s$  = target area

This equation (2.3) can be simplified under the following three assumptions (Jelalian, 1992):

- The target area ( $A_s$ ) has circular shape and the entire footprint is reflected on one surface (extended target). Hence, it can be defined as the function of range ( $R$ ) and laser beam width ( $\beta_t$ ).

$$A_s = \frac{\pi R^2 \beta_t^2}{4} \quad (2.4)$$

- The scattering solid angle ( $\Omega$ ) of the whole targets is  $\pi$  steradians. It is  $2\pi$  in radian unit ( $\Omega = 2\pi$ ).
- The whole surface reflects laser shots by Lambertian cosine law, which means that intensity observed from a surface is directly proportional to the cosine of the angle ( $\alpha$ ) between the observer's viewing direction and the surface normal.

$$\sigma' = \sigma \cdot \cos \alpha \quad (2.5)$$

Substituting these into equation (2.3), the target cross-section ( $\sigma$ ) is expressed into the following equation.

$$\sigma = \pi \rho \cdot R^2 \cdot \beta_t^2 \cdot \cos \alpha \quad (2.6)$$

The equation (2.2) also contains sensor-related factors such as  $P_t$ ,  $D_r$  and  $\eta_{sys}$ . However, most sensor systems do not record or provide these sensor-related factors. Hence, the prerequisite is required that one sensor system transmits pulses to have constant signal power. Furthermore, as laser scanning data are collected by one sensor system during one campaign, all of sensor-related factors can be set to be a constant factor within a certain campaign.

The atmospheric factor ( $\eta_{atm}$ ) stands for the loss of energy mainly caused from scattering and absorption while a pulse travels between sensor and target. Under the assumption that the atmospheric condition between sensor and target is constant, it is expressed to the function of the range (2.7), where air attenuation coefficient ( $a$ ) can be derived by the estimation by exponential function of range and intensity value (Coren and Sterzai, 2006) or the approximate value by the former physical experiment (Höfle and Pfeifer, 2007). For example, for horizontal propagation air attenuation ( $a$ ) has the average value of 0.2 dB/km in extremely clear atmospheric conditions and 3.9 dB/km in haze conditions. The average vertical attenuation increases with higher altitudes. It has 0.22, 0.17 and 0.14 dB/ km for flying height of 1000m, 2000m and 3000m respectively in mid-latitude summer with a visibility of 25km.

$$\eta_{atm} = 10^{-2Ra/10000} \quad (2.7)$$

Where,  $a$  = atmospheric attenuation coefficient (dB/ km)  
 10000 originates from  $a$  given in dB/ km

As a result, replacing parameters in equation (2.2) into derived equations, it is simplified into the following equation.

$$P_r = \frac{\rho}{R^2} 10^{-2Ra/10000} \cdot \cos \alpha \cdot C \quad (2.8)$$

$$\rho \propto R^2 \cdot 10^{2Ra/10000} \cdot \cos^{-1} \alpha \cdot C \quad (2.9)$$

Where,  $C$  = constant factor on sensor system ( $P_t$ ,  $D_r$  and  $\eta_{sys}$ )

$\rho$  = corrected value of intensity to include surface property

This equation was tested by Coren and Sterzai (2006), where the range was calculated by the distance from sensor to target and the incidence angle was replaced by scan angle. As presented in Figure 2-1, the uncorrected data have the variation of intensity values between strips whereas the variation is significantly reduced in the corrected result. Also shown is better contrast between different surfaces. Moreover, it was noticed that this correction can be considered as a potential aid to enhance the accuracy of laser data classification.

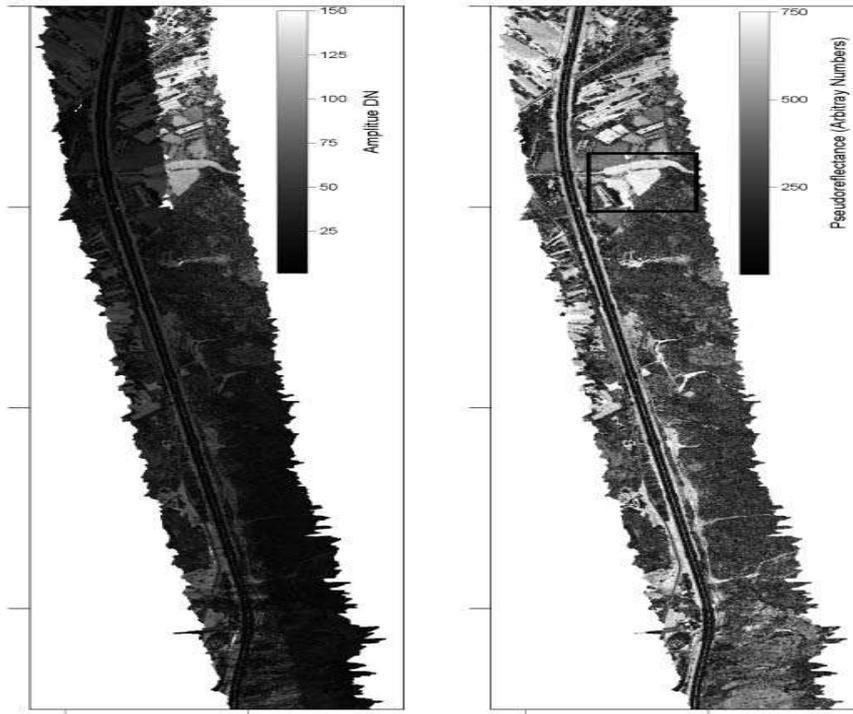


Figure 2-1: Measured (left) and corrected (right) result (Coren and Sterzai, 2006)

### 2.2.2. Method based on the empirical model

This method is based on the equation derived from data itself rather than physical law or equations. It requires the fundamental assumption that intensity values from the same surface are identical. The researches based on empirical model were processed by the following procedures:

- i. Selecting the suitable equation on the relation between the received intensity and influencing factors
- ii. Inspecting the change patterns of the recorded intensity value for influencing factors through sampling from homogeneous surface
- iii. Defining the best-fit model based on the selected equation and estimated parameters for the inspected patterns and defining the final equation.

As introduced in section 2.1, several studies on intensity correction have been done by this method. In this section, two different attempts in ALS are introduced.

### 2.2.2.1. Empirical model based on range only (Höfle and Pfeifer, 2007)

For this approach it is assumed that the received intensity values are proportional to the ground surface reflection and depend on the ranges via monotonic functions. Under this assumption, inversely quadratic (2.10) and Taylor expansion (2.11) function were selected to express the relation between the received intensity and ranges.

$$f(r) = \frac{1}{ar^2 + br + (1 - 1000^2 \cdot a - 1000 \cdot b)} \quad (2.10)$$

$$f(r) = (1 - 1000 \cdot c_1 - 1000^2 \cdot c_2 - \dots) + c_1 r + c_2 r^2 + \dots \quad (2.11)$$

By the parameters in equations, all physical effects were simply included such as sensor-related and atmospheric parameters. It did not take into account the effect of incidence angle as it is applied in ALS data, which has the restricted scan angle of the maximum  $\pm 30$  degrees (FLI-MAP 400) and the range is relatively longer so that the effect of incidence angle by surface orientation is negligible. In order to estimate these parameters, sets of points from at least three notably different ranges were sampled in cells with homogeneous intensity. After the estimation of parameters, the equation was defined and it was applied to the whole dataset.

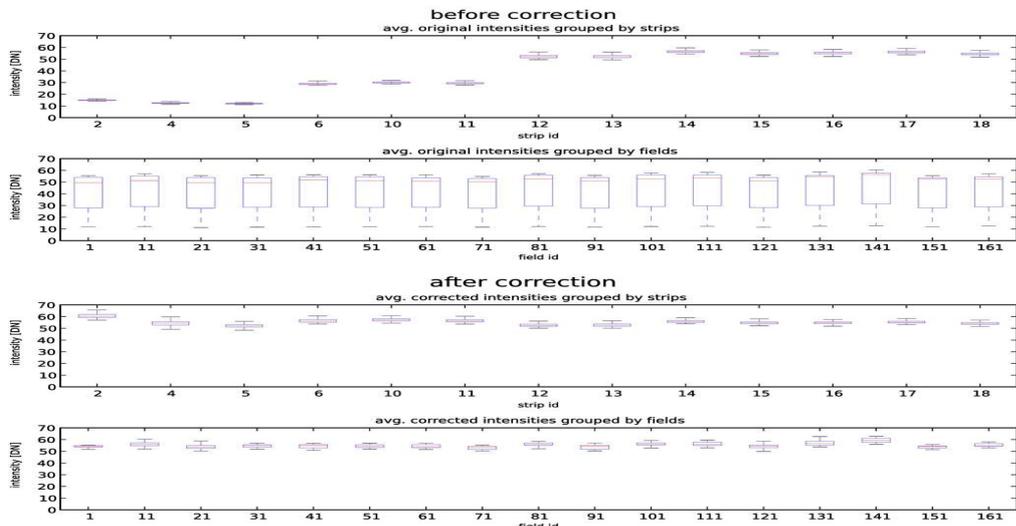


Figure 2-2: Recorded and empirically corrected intensities (Höfle and Pfeifer, 2007)

In Figure 2-2, a field corresponds to a set of points from a homogeneous area with different ranges. The corrected result showed the significant decrease of variation between strips with different flying height even though it does not take into account the incidence angle. As a result, this paper noticed the usability of empirical model for intensity data correction.

#### **2.2.2.2. Adjusted radar range equation based on range and incidence angle (Jutzi and Gross, 2009)**

In order to improve the correction of laser scanning data, the authors basically employed the simplified radar range equation. The difference to the approach by the radar range equation is the attempt to modify the contribution of each influencing factor by the change of constant values that have been derived from the theoretical equations.

$$I_A = I \cdot R^a \cdot 10^{2bR} \cdot \cos^c(\theta) \cdot 10^d \quad (2.12)$$

Where,  $I$  and  $I_A$  = Measured and corrected intensity

$a$  = beam divergence

$2bR$  = air attenuation by two way propagation

$c$  = type of reflectivity (effect of incidence angle)

$d$  = constant for normalization to 1

The equation (2.12) shows that it uses four parameters ( $a$ ,  $b$ ,  $c$ , and  $d$ ) instead of the values by the computational simplifying procedure. In section 2.2.1 it has been derived as  $a = 2$ ,  $b = 0.2 \sim 0.95$  dB/km,  $c = -1.0$ , and  $d$  = system-related constant factor. Here, those parameters are estimated through logarithmic linear least square sum by sets of points sampled from homogeneous surface with different ranges and incidence angles.

When this approach was tested in ALS dataset, the authors employed the planar roof surfaces to estimate the surface orientation (surface normal). It was calculated two steps: a) the selection of roof surfaces and b) the estimation of surface normal made by 3D neighbours of each point within a belonging roof. Firstly, the points in roof surface were carefully sampled through the segmentation by the covariance matrix, the corresponding eigenvalue and eigenvector of small 3D neighbours of each point (Gross and Thoennessen, 2006). Exterior points of roof surface were filtered out by planarity based on descend-sorted eigenvalues of the covariance matrix within each planar roof segment (West et al., 2004). After this process, the surface normal of each point was estimated by the neighbours within a certain distance. For this dataset, the constant parameters were estimated as  $a = 2.08$ ,  $b = 0.52$  dB/km,  $c = -0.60$ , and  $d = -21.42$ , which were slightly different with theoretical equation (2.9).

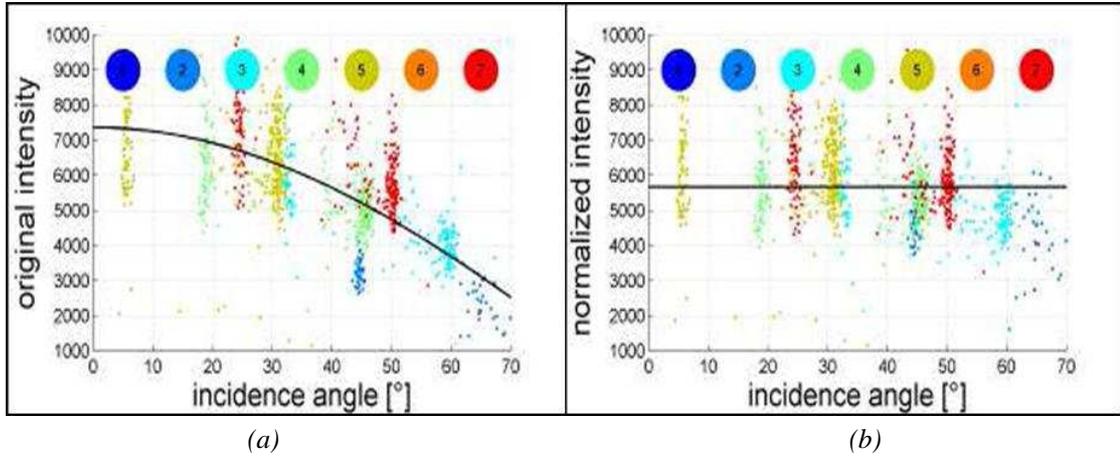


Figure 2-3: Intensities for incidence angle (Jutzi and Gross, 2009)

In Figure 2-3, the intensity values of all points inside each sample were coloured by the corresponding strip. It was noted that the original intensities for incidence angle have the trend of the approximate cosine curve (a) whereas the corrected intensities become constant (b).

Furthermore, the authors tried to improve the correction model by the empirical Phong surface model considering diffuse and specular scattering characteristics of the surface. The Phong surface model contains the effect by specular reflection, ambient lightning as well as diffuse scattering that has been dealt with by Lambertian cosine law (Phong, 1975). Ignoring ambient lightning, the equation (2.12) is extended to the following equation (2.13).

$$I_A = I \cdot R^a \cdot 10^{2bR} \cdot 10^d / ((1 - k_s) \cdot \cos(\theta) + k_s \cdot \cos^n(2\theta)) \quad (2.13)$$

Where,  $k_s$  = weighting for the specular reflection

$1 - k_s$  = weighting for the diffuse reflection

In the ALS experiment this extended approach showed only a small improvement. The authors concluded that the result derived by Lambertian model (equation (2.12)) seems to be sufficient as the improvement might result from the backscattering characteristic of the sampled roof material.

### 2.3. Experiments by reference targets

As shown in section 2.2, the results showed the improvement on the variation of intensity data. However, its application is still limited due to the difficulty to formulate data reflected from natural surfaces with one universal equation such as the radar range equation and empirical model. Therefore, the understanding of performance properties of laser scanning in the field is critical to deal with intensity data. With this motivation, several experiments by reference target have been done to inspect the change of laser shots flying in the field and two experiments useful for MLS intensity data are introduced in this section.

### 2.3.1. Influence by incidence angle (Kukko et al., 2008)

As it has been realized that incidence angle has an impact on the reduction of signal power, the incidence angle has been dealt with by Lambertian cosine law or by Phong surface model. However, those physical models tested in laboratory conditions were not enough to describe the exact phenomena in the field. Moreover, unreasonable assumptions were required, e.g. all surfaces have Lambertian scattering characteristics. Thus, this experiment was done according to the requirement for a systematic set of data on the effect of the angle of incidence. In the meantime, surface roughness and brightness were considered.

The experiment was done by two laser scanning sensors: a) FARO LS 880HE80 TLS operating at 785 nm and 2) TopEye MKII operating at 1064 nm. A set of five targets with calibrated reflectance of 8%, 26%, 50%, 70 % and 99% and a set of five targets with sandblasting of two gain sizes (0.1-0.6 mm and 0.5-1.2 mm), black gabro gravel, crushed redbrick, and light expanded clay aggregate (LECA) were used as the reference targets respectively for brightness and roughness.

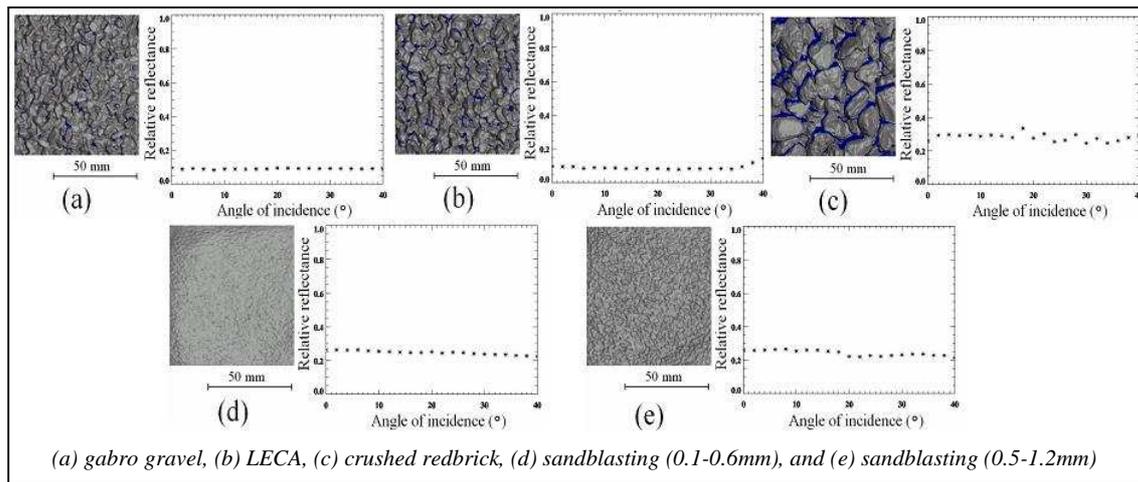


Figure 2-4: The effect of incidence angle for surface roughness (Kukko et al., 2008)

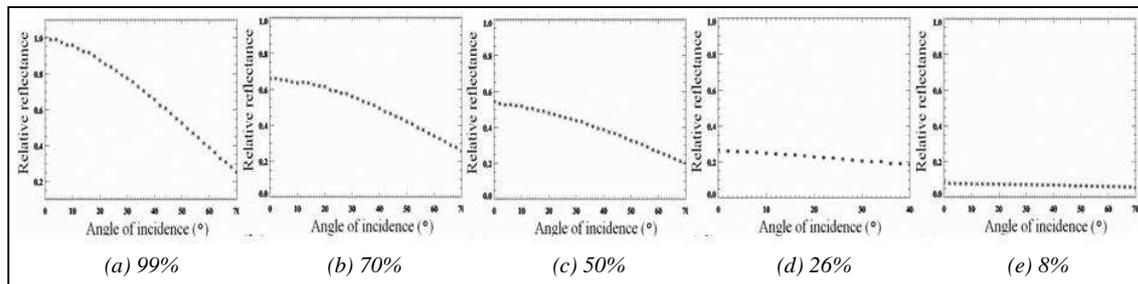


Figure 2-5: The effect of incidence angle for surface brightness (Kukko et al., 2008)

In the first test done by TLS, it was noted that the decrease in intensity is stronger at targets with finer gain size and the variation of intensity is stronger at rough targets due to uneven gain size and sharp edges (Figure 2-4). As shown in Figure 2-5, it was noticed that the decrease of intensity in the bright target is significant at incidence angles of more than 20°. On the other hand, the decrease of intensity in most targets except for the brightest is practically negligible in incidence angles of less than 20°.

Type		TopEye ALS (1064 nm)			FARO TLS (785 nm)		
Incidence angle		12 °	18 °	20 °	12 °	18 °	20 °
Bright- ness	50%	0.68	0.68	0.72	0.79	0.79	0.78
	20%	0.36	0.37	0.38	0.24	0.24	0.23
	10%	0.12	0.12	--	0.12	0.12	0.12

Table 2-1: Relative intensity value in ALS and TLS in low angles of incidence (Kukko et al., 2008)

In the second test, TLS data were compared with ALS data taken from 300 m of flying altitude. When the values were expressed in relation to the 70% brightness target with the same incidence angle, it was noticed that relative results are well produced in smaller incidence angles (Table 2-1). From above two tests, the authors concluded that the incidence angle must be considered in the interpretation of laser scanning data collected in the area with steep slopes or vertical surfaces and the brightness causes more decrease of intensity in larger incidence angles.

### 2.3.2. Influence by sensor own property based on range (Pfeifer et al., 2008)

As the correction methods work successfully in ALS data, this experiment was conducted for the correction method of TLS data. A Riegl LMS-Z420i (Riegl) and an Optech ILRIS 3D (Optech) were employed that operate respectively at the wavelength of approximately 1550 nm and 1540 nm. The reference targets consisted of a set of six targets with reflectivity of 5%, 20%, 40%, 60%, 80%, and 99% that reflect according to the Lambertian cosine law.

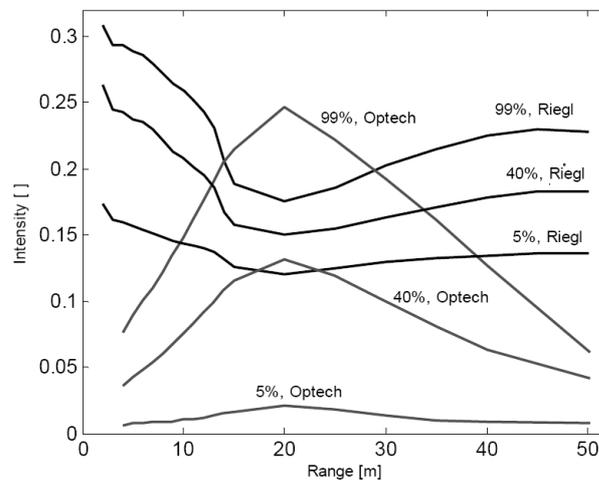


Figure 2-6: Mean intensities for Riegl and Optech at different distances (Pfeifer et al., 2008)

As the first test was performed to inspect the change of intensity values for the distance in each sensor, the reference targets were located at different distances from the sensor: up to 15 m in 1 m-steps, then up to the maximum length of the laboratory (50 m) in 5 m steps. As the calculated range, incidence angle and measured intensities within a target varied, the variation was dealt with by the average. As shown in Figure 2-6, in the test for distance the measurements in both sensors were completely different to the expectation by the radar range equation. In the second test the intensities for different incidence angles were measured, where the reference targets were placed at approximately 15 m from the sensor and rotated in 9° steps from 0° to 72°.

From the results by two different TLS sensors, the authors derived the conclusion that radar range equation has obvious limitations in TLS and proposed the approaches by best-fit models as possible solutions for correction.

## **2.4. Summary**

As introduced in this chapter, several researches have been contributed in correction for intensity data. They can be classified into two categories based on their basic principles. One of them is based on the theoretical radar range equation and Lambertian surface law. Another is based on the empirical models derived from data.

The theoretical model-based approach has focused on reasonable simplification of the radar range equation as the first step. In a simplification procedure most of system-related factors such as receiver aperture diameter, system transmission factor and beam width are ignored as a constant factor except for air attenuation factor. The air attenuation is assigned by estimation from data (Coren and Sterzai, 2006) or former physical result (Höfle and Pfeifer, 2007). In ALS data the corrected results showed the improvement for consistency of intensities in homogeneous surface.

As experiments by the reference target show the results which do not meet theoretical equations, the second approach was tried. The empirical model-based approach does not consider theoretical bases and focuses on a standard procedure of selection of proper model, model fitting by data and diagnostics. The corrected results also showed the improvement in both ALS and TLS. From the experiment it was turned out that intensity data in near ranges are changed differently to the expected by radar range equation. Therefore, it was demonstrated that the radar range equation can apply to a certain range interval of TLS data (Pfeifer et al., 2008).

However, there is still a need for the improvement towards developing correction models for MLS data. Based on the existing methods, the behaviour of intensity needs to be inspected with respect to range and incidence angle in MLS. In addition, the adjustment between sensors has to be studied that is not mentioned in previous research since MLS usually uses more than two sensors simultaneously.

## 3. Proposed Method

### 3.1. Overview

As shortly introduced in chapter 1, intensity data will be corrected by two approaches: a) by theoretical model (radar range equation) and b) by empirical model. As shown in Figure 3-1, outliers and multi-echoes are filtered out from the point clouds during the pre-processing step. For the remaining points range and incidence angle are computed with the help of the GPS trajectory and offset data. As it is known that the range and incidence angle are the main influencing factors on intensities, they are used in the correction procedure of two different approaches. Finally, each output is reproduced by the proposed models. Figure 3-1 shows the main steps in this study. Each step will be explained in detail in the following four sections.

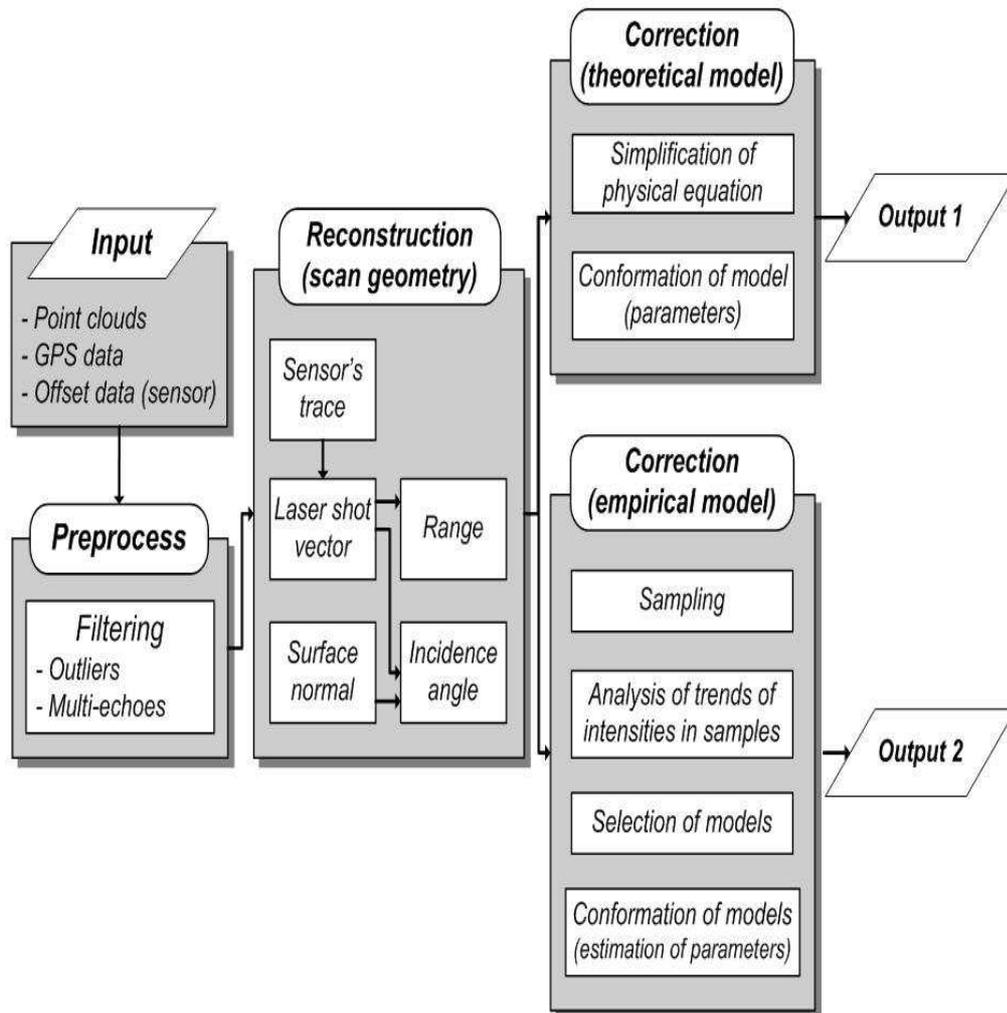


Figure 3-1: Workflow of proposed method

### 3.2. Pre-processing

Compared to images, intensity data of a laser scanner have been known to be relatively noisy. It has been identified that noise is caused mainly from under-sampling by low point densities in ALS data. They are considered as other reasons that most laser scanners measure intensity data by quantifying the intensity value at a certain time rather than integrating values over a small period around the specified time, and one emitted echo is recorded as multi-echo when it hits multiple objects (Vosselman, 2002). In order to remove this sort of intensity noise, the recent laser scanning technology has been improved and can acquire extremely high point densities. However, the effect of pepper and salt cannot be completely removed due to other noise characteristics and deteriorates the quality and accuracy of the intensity-based application. Therefore, pre-processing as the first step is dedicated to determine input data. It focuses on filtering out multi-echoes and outliers, which cannot be used in the correction of intensity data.

- **Multi-echoes:** If a laser pulse hits partly an obstacle, it can be split into more than two pulses that are called multi-echo. They are realized as noises in intensity data as their measurements are different to single echoes. However, it is not possible to estimate how much proportion of a laser shot is divided into each multi-echo with the data by the sensor system. Therefore, this property makes it impossible to correct intensity measurements of multi-echoes. Multi-echoes will be excluded and not be considered for the correction.
- **Outliers (specular reflection):** as mentioned earlier, the measurements of intensity data are quantified at a certain time estimated as the peak of waveform. Laser pulses from specular surfaces are measured as much higher values than pulses from normal surfaces since they have more sharp waveforms. Moreover, as specular reflections have high intensity in nadir or low intensity in off-nadir, they can be considered as outliers. Therefore, they can be filtered by the analysis of their histogram. As a laser scanner does not collect low intensities by the device threshold for data acquisition, only high intensities in nadir need to be filtered out.

### 3.3. Reconstruction of scan geometry

This step is dedicated to obtain the required data for correction of intensity data, or laser shot vector (range and incidence angle). It is processed in the following workflow (Figure 3-2). The required input data are point clouds, corresponding GPS trajectory and the offset data of sensor device.

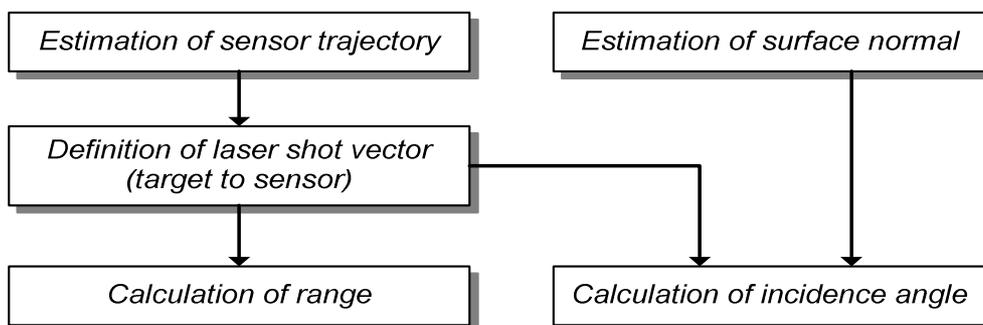


Figure 3-2: Workflow for reconstruction of scan geometry

### 3.3.1. Range

In principal, range can be calculated by the length of the laser shot vector from target to sensor. According to Figure 3-2, the sensor trajectory is estimated from GPS data and offset data which describe the distance from GPS to sensor. The GPS data used in this study have been already integrated by GPS and IMU measurements. As the provided offset data is defined in sensor's own coordinate system, the corresponding sensor point of each GPS point can be computed by rotating the offset distance into the coordinate system of the point cloud and subtracting it from the GPS point.

As depicted in Figure 3-3, the vector from target to sensor can be defined with two points (each echo and the corresponding sensor position). Here, the corresponding sensor position of each echo can be found using the time stamp. However, as the scan shot frequency is much higher than the GPS frequency, the corresponding time stamps of the echoes in the point clouds cannot be found from GPS data directly. In order to reduce the error in this process, the sensor position is estimated between the nearest two sensor positions using the time stamps by linear interpolation. The estimation by linear interpolation can be considered to be reasonable for MLS data since the interval between GPS positions is small. For example, if a vehicle with GPS of 200 Hz drives about 60 km/h, the distance between neighbours is at most 8.33 cm. As shown in Figure 3-3, the zoomed sensor's trace is a line of points that is approximately close to a straight line. It is another proof that linear interpolation is reasonable method. Finally, the range of each point is reproduced by the length of the vector pointing from each target and the interpolated sensor point. Once this is done, points with the range above a specified maximum range are removed from dataset since they can be considered as noise by mis-measurement.

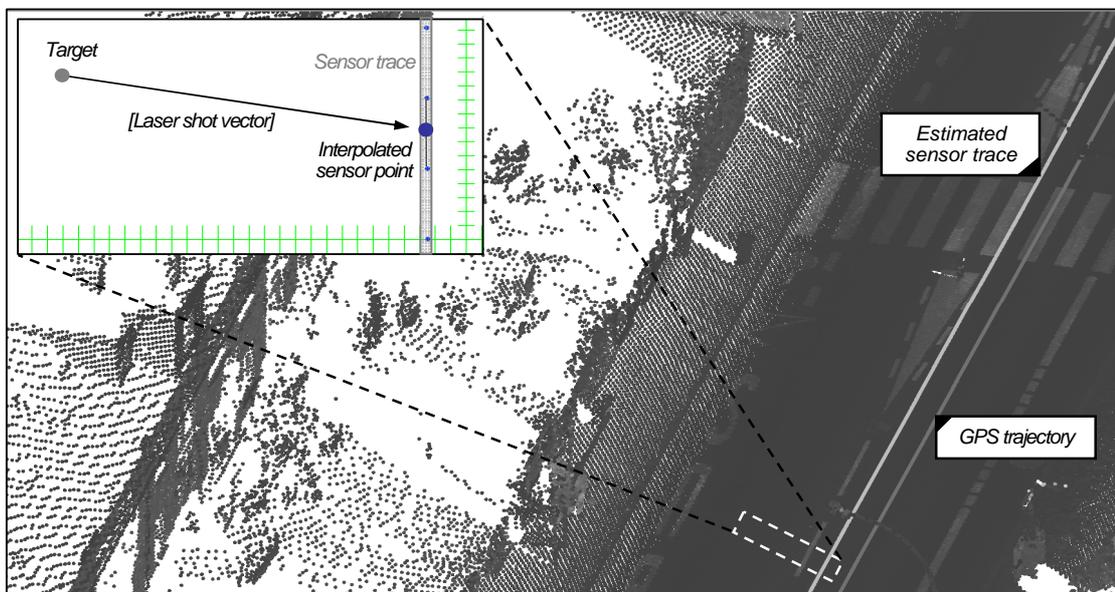


Figure 3-3: Description of range calculation (points were represented by intensity value.)

Figure 3-3 describes the procedure for the calculation of range data, where tagged two gray lines mean sensor and GPS trajectories and the unit of a scale bar in the zoomed image is 1 cm.

### 3.3.2. Surface normal and Incidence angle

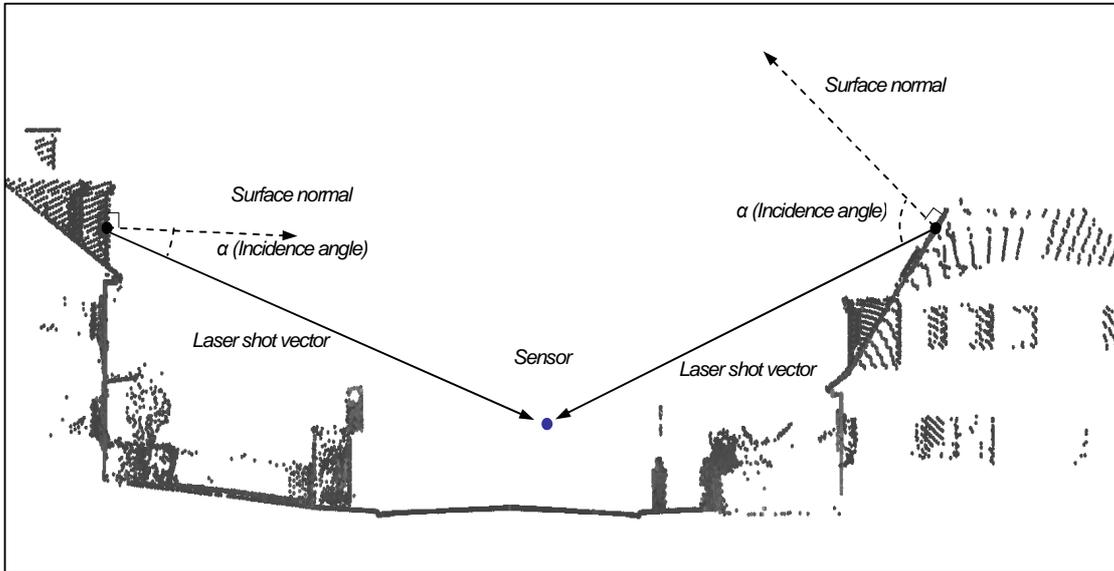


Figure 3-4: Definition of incidence angle (profile view - represented by intensity value)

As Figure 3-4 shows, the angle of incidence can be yielded by the angle enclosed between surface normal and laser shot vector produced during the range calculation (section 3.3.1). Thus, the surface normal of each point is required in this step. It can be derived from the  $k$  nearest points found by KNN ( $k$ -nearest neighbours) algorithm that uses data structure named kd-tree. However, the accuracy on the estimation of surface normal has an unavoidable problem in calculation, e.g. in vegetation, rough surfaces and round or sharp artificial objects. Due to the positioning errors caused from GPS and IMU and ranging errors, the estimation is still erroneous even in simple flat or smooth surfaces. Thus, the use of the approximately estimated surface normal is inevitable. The estimation can be categorized in two ways according to where to find  $k$  nearest neighbours: a) within each segment and b) within whole dataset.

Comparing the two methods, the first one to take neighbours only within the corresponding segment is effective to reduce the variation of surface normal in planer segments with large number of points and area. This method works better at corners to meet planar surfaces such as facades of buildings and curb stones of roads. It was used by Jutzi and Gross, 2009 that defined the orientation of each point on sampled roof surfaces. However, the result of this method is dependent on segmentation methods, requires more computation and is especially problematic in segments with small area and low point density. Since the estimation by small number of points and area cannot overcome the intrinsic error in point clouds, it makes estimations more erroneous. On the other hands, the second method to take neighbours within whole point clouds is proper to smooth surfaces and requires less computation time for estimation. However, it considers surfaces to be smooth and it does not work on edges and artificial objects.

Based on above analysis, the second method is applied in this research since most points of MLS are collected from planer surfaces such as roads and facades of buildings that are facing towards road side. Furthermore, it has to be mentioned that the correction method can be employed only in relatively smooth surfaces in order to solve the problem of estimation error in sharp artificial objects. Also, this

needs to be taken into account in upcoming processes such as sampling scheme and evaluation of models. For example, sampling has to be done in only smooth surfaces and the evaluation needs to be based on smooth surfaces excluding the performance at sharp corners.

After the nearest neighbour searching, the estimation of the surface normal is performed which requires expensive computation. Among several developed algorithms, the robust least square sum can be considered, however, it has been noticed that this approach can be applied only in the case that z-component is functionally dependent on x- and y- components (Eberly, 2008). As 3D geometry in laser scanning data does not meet the assumption, the estimation by planer regression is used in this study. The planer regression has been proven as the simple and effective method to calculate representative values for surface orientation from sets of points (Fernández, 2005). For sets of neighbours constituted vertically or horizontally, the following two equations are applied depending on the variance of each coordinate (Table 3-1). It is considered that the first equation can handle vertical sets of points such as building walls and the second equation can be applied to horizontal sets of points such as road surfaces.

<i>Coordinate with min. variance</i>	<i>X or Y (vertical set of points)</i>
<i>Equation</i>	$(x_i - \bar{x}) = B \cdot (y_i - \bar{y}) + C \cdot (z_i - \bar{z})$
<i>Estimation equations</i>	$\sum((x_i - \bar{x})(y_i - \bar{y})) = B \cdot \sum(y_i - \bar{y})^2 + C \cdot \sum((y_i - \bar{y})(z_i - \bar{z}))$ $\sum((x_i - \bar{x})(z_i - \bar{z})) = B \cdot \sum((y_i - \bar{y})(z_i - \bar{z})) + C \cdot \sum(z_i - \bar{z})^2$
<i>Surface normal</i>	$(-1.0, B, C)$
<i>Coordinate with min. variance</i>	<i>Z (horizontal set of points)</i>
<i>Equation</i>	$(z_i - \bar{z}) = A \cdot (x_i - \bar{x}) + B \cdot (y_i - \bar{y})$
<i>Estimation equations</i>	$\sum((z_i - \bar{z})(x_i - \bar{x})) = A \cdot \sum(x_i - \bar{x})^2 + B \cdot \sum((x_i - \bar{x})(y_i - \bar{y}))$ $\sum((z_i - \bar{z})(y_i - \bar{y})) = A \cdot \sum((y_i - \bar{y})(x_i - \bar{x})) + B \cdot \sum(y_i - \bar{y})^2$
<i>Surface normal</i>	$(A, B, -1.0)$

Table 3-1: Equations and surface normal adopted according to the arrangement of neighborhood

### 3.4. Data correction based on radar range equation

This step is devoted to investigate whether the theoretical model can be used in the correction of MLS data. Thus, the simplified radar range equation ( $\rho \propto R^2 \cdot 10^{2Ra/10000} \cdot \cos^{-1} \alpha \cdot C$ , equation (2.9)) is directly used which is based on the radar range equation (Jelalian, 1992) and Lambertian cosine law. Even if it showed improved results in ALS data (Coren and Sterzai, 2006; Höfle and Pfeifer, 2007), a test by TLS sensors showed that the intensity values in near ranges have the different pattern than the theoretical model (Pfeifer et al., 2008). Thus, it can be predicted that the quality of result would be

low. However, it is still valuable to apply to MLS data as it has not been tested and can help to understand the behaviour of intensity data in dataset.

### 3.4.1. Definition of parameters

As equation (2.9) includes two parameters related on constant factor ( $C$ ) and air attenuation ( $a$ ), they need to be defined for further calculation. Firstly, constant factor is associated to sensor system which includes the system factor, transmitted signal power, and aperture diameter of receiver. They are not usually provided by sensor provider. Moreover, if they are defined as constant values, they are meaningless for the relative correction and affect only to compress or stretch the extent of intensity data. Thus, it is defined by trial-and-error only for proper visualization of data. Secondly, the air attenuation coefficient has the main contribution from the absorption and scattering of transmitted laser shots by aerosols and molecule in the atmosphere depending on its wavelength. Therefore, the coefficient can be derived as the function of visibility and wavelength (Kim et al., 2001). Table 3-2 shows the coefficient for the wavelength 1,550 nm, which is close to the wavelength of LYNX sensor of Optech (1,500 nm). It is used for air attenuation coefficient in radar range equation. However, since MLS normally takes point clouds from the range less than 50 m, it can be anticipated that air attenuation factor will not have much influence in the correction procedure of intensity value.

Visibility (km)	23	10	4	2	1	0.5	0.2	0.05
Coefficient (dB/km)	0.2	0.4	2	4	10	34	85	340
Weather	Clear		Haze			Fog		

Table 3-2: A table of atmospheric losses for the wavelength of 1,550 nm (Kim et al., 2001)

### 3.4.2. Simulation by simplified radar range equation

As the correction model is defined, the behaviour of intensities can be simulated to predict the effect of influencing factors, or range and incidence angle. In this simulation all assumptions in the computational simplification procedure are valid (section 2.2.1) and it is additionally assumed that the simulation is for intensities reflected from the surface with homogeneous reflectivity. For two parameters, the air attenuation coefficient ( $a$ ) was supposed to be 0.2 dB/km in clear weather condition and the constant factor ( $C$ ) was set up to have intensity of 1.0 in 5 m of range and 0° of incidence angle ( $C = 25$ ), thus the intensity values are hereafter expressed in the relative value on it.

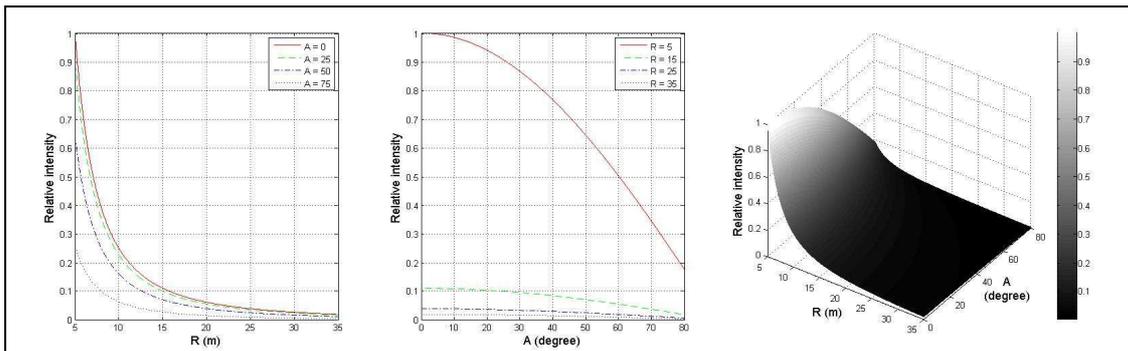


Figure 3-5: Simulation of intensity data by radar range equation ( $R$ : range,  $A$ : incidence angle)

In Figure 3-5 the intensity is simulated in 5 ~ 35 m of range and 0 ~ 80° of incidence angle as intensities become close to 0 or infinite value in the exterior of these scopes. These results point out that the range influences more strongly to decreases of intensity than what the incidence angle does. The incidence angle has more effect in near ranges where pulses maintain higher intensity values.

### 3.5. Data correction based on empirical models

Aside from the theoretical model base approach, recently the possibility for correction of intensity data by empirical models has been investigated (Höfle and Pfeifer, 2007; Jaakkola et al., 2008; Jutzi and Gross, 2009; Pfeifer et al., 2008). In order to find the general model, the experiment has been performed by examining reference targets. However, intensities are sensitive to the circumstance such as weather, temperature and air pressure. It is still difficult to find one general model or equation even though it is modelled by the measurements of reference targets. In the same token, the problem on stability of a correction model depending on temporal changes has been suggested (Pfeifer et al., 2008). In other words, the correction model can require differently defined models depending on the circumstances at the time of acquisition. Thus, it can be proposed that intensity needs to be corrected by the model defined by data in a certain survey. In this study it is carried out to analyze the plots by samples and define the correction models by analyzed results.

#### 3.5.1. Analysis of uncorrected intensities

##### *Sampling*

In the analysis and estimation by sampling, sampling scheme is a crucial part since samples work as representative of the entire dataset and poor sampling scheme can result into significant biases and over or under estimation of results (Congalton, 1991).

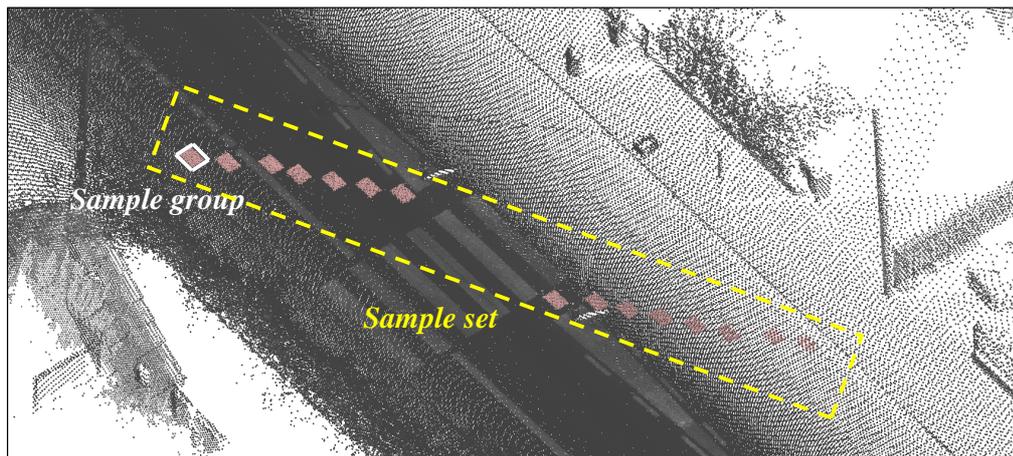


Figure 3-6: Description of used terms in sampling procedure

The sampling is performed by following five sampling scheme. As shown in Figure 3-6, a sample set consists of sample groups and each sample group includes points in a rectangular patch on a homogeneous surface.

- Manual sampling: Recently, as a laser scanning technique has been accepted as the most reliable 3D data, there is no geo-referenced 3D data available that have more precise

information on surface types of surveyed region. Thus, manual sampling is inevitable and it is done by the similar size of patches, which form sample groups.

- **Stratified sampling:** To define the correction model, points are sampled from several homogeneous surfaces which are expected to have similar reflectivity. Each surface type makes a sample set that consists of a number of sample groups.
- **From regions with various range and incidence angle:** In order to keep errors of the defined model minimal, sample groups have to cover all situations of range and incidence angle which might occur. Also, important is to minimize variance of range and incidence angle within a sample group since a set of representative values is used in a modelling procedure. Therefore, sampling patches are created in parallel with GPS trajectory. Also, the patch size is decided considering a local point density.
- **From smooth planer surface:** As mentioned, the estimation of surface normal has more error on rough surfaces and the edge of artificial objects. This kind of surfaces makes the estimation of incidence angles wrong. Moreover, this results into incorrect modelling. Therefore, samples are selected from smooth surfaces excluding them.
- **Removing of samples with low intensities:** Despite the automatic removal of low intensities by the device, there is another problem in low intensity data. Since intensities are of nominal values recorded by lookup table, arithmetic calculations cannot be applied. More serious arithmetic calculation is shown in lower intensities. For example, when intensities would be supposed to be recorded by Table 3-3, real signal power of 0.1 would never be twice of 2.0 even if it would be arithmetically computed to be twice by nominal intensity values. However, it is acceptable in larger intensities as signal power of 20.0 is close to twice of 9.1. Thus, samples with low intensities have to be excluded in modelling procedure. The applied filtering threshold depends on dataset property such as surveyed area and point density. In addition, high noises do not need to be considered in this process, as they have already been removed in previous process.

<i>Intensity value</i>	<i>1</i>	<i>2</i>	<i>...</i>	<i>10</i>	<i>...</i>	<i>20</i>	<i>...</i>
<i>Signal power</i>	<i>0.1 ~ 1.0</i>	<i>1.0 ~ 2.0</i>	<i>...</i>	<i>9.0 ~ 10.0</i>	<i>...</i>	<i>19.0 ~ 20.0</i>	<i>...</i>

*Table 3-3: Generic example of lookup table of intensity with respect to signal power*

### ***Analysis of samples***

After careful sampling, it is focused to analyze the changing pattern of intensities for range and incidence angle. The analysis is separately done by each sensor as both sensors have different system properties. However, some problems can still arise in this analysis due to performance properties of MLS. Firstly, it can be challenging to find a certain trend in MLS data due to variation of intensities reflected from natural surfaces. Moreover, since MLS collects point clouds from surrounding area of roads, most geometry of point clouds has similar distribution. This distribution is unavoidable in MLS due to mapping along the road. For instance, samples from paved roads by asphalt occur mostly with smaller incidence angles in near range and larger incidence angle in far range. Building walls and roofs do not occur in near range. This property makes it difficult to inspect the distribution patterns by

various ranges under the same incidence angle and vice versa. Namely, it is not possible to examine variables independently from each other.

For the problem on variation of intensities the analysis by box plot can be proposed that has been used in laser scanning research (e.g. Höfle and Pfeifer (2007) and Jutzi and Gross (2009)). As each sample group is shown by median, quartiles, max/ min and range of the distribution, it makes the comparison between sample groups more clear. A possible solution for the similar distribution patterns is to inspect two variables simultaneously rather than independently. However, this will affect to model selection, that is, empirical models separated by each variable cannot be expected.

### 3.5.2. Adjustment of the radar range equation

As addressed in section 2.2.2.2, this approach was tested in ALS intensity data (Jutzi and Gross, 2009). The same approach is also chosen to be applied in MLS intensity data in this research.

$$\hat{\rho} = I \cdot R^a \cdot 10^{2bR} \cdot \cos^c(\theta) \cdot 10^d \quad (3.1)$$

Parameters (a, b, c, and d) from equation (3.1) are estimated by logarithmic linear regression approach with the following equations (3.2) and (3.3). The relative reflectivity ( $\hat{\rho}$ ) is estimated considering the brightness of sampled homogeneous surfaces. Practically, it can be defined as the intensity values with the similar range and incidence angle in sample sets. As these values are identically assigned to both sensors, it can be expected that the same surface type will be adjusted to similar intensity in both sensors. However, since the brightness of each homogeneous surface cannot be accurately quantified in natural surfaces, the parameters are separately estimated by each sample set and compared with each other.

$$\log I_A = \log I + a \cdot \log R + 2bR + c \cdot \log(\cos \theta) + d \quad (3.2)$$

$$\begin{bmatrix} \log R_1 & 2R_1 & \log(\cos \theta_1) & 1 \\ \log R_2 & 2R_2 & \log(\cos \theta_2) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \log R_n & 2R_n & \log(\cos \theta_n) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \log I_A - \log I_1 \\ \log I_A - \log I_2 \\ \vdots \\ \log I_A - \log I_n \end{bmatrix} \quad (3.3)$$

Where, n = the number of points in each sample

Nevertheless, there is a problem in this estimation due to outliers and different size of sample groups within a sample set. Since sample groups are chosen by rectangular patches with the similar size, the number of points in sample groups is different depending on local point density. This results to over-estimation in a sample group with large size. In order to reduce this effect, the median values of variables in each sample group are used in estimation as representatives. Therefore, the following equation (3.4) can be populated.

$$\begin{bmatrix} \log R_{M1} & 2R_{M1} & \log(\cos \theta_{M1}) & 1 \\ \log R_{M2} & 2R_{M2} & \log(\cos \theta_{M2}) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \log R_{Mk} & 2R_{Mk} & \log(\cos \theta_{Mk}) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \log I_A - \log I_{M1} \\ \log I_A - \log I_{M2} \\ \vdots \\ \log I_A - \log I_{Mk} \end{bmatrix} \quad (3.4)$$

Where,  $R_M$ ,  $\theta_M$ , and  $I_M$  = median value of each variable

k = the number of sample groups in each sample set

After the estimation of parameters, the correction model can be defined. However, the problem of extrapolation must be considered additionally. The problem has been studied in statistics and it was concluded that the standard procedures of model selection, model fitting and diagnostics can be inappropriate in extrapolation problems (Lavine, 1991). However, it is not possible to find sample sets to cover the entire extent of range and incidence angle due to the mentioned similar surrounding distribution of surface types in the point cloud. Therefore, the correction model is broken into more than two parts in order to perform the correction of whole intensities by interpolation. Finally, corrected dataset is reproduced by this model.

### 3.5.3. Data fitting

It has been revealed that radar range equation is not appropriate in TLS to express the change pattern of intensities in near range. At the same time, it has been noticed that the change of intensities in TLS can be also expressed by new models that are not associated with the theoretical model (Pfeifer et al., 2008). Therefore, this step is conducted by the standard procedure on empirical model of model selection, model fitting and diagnostics.

Firstly, the change of intensities with respect to range and incidence angle is re-inspected by box plot of each sample set. However, as addressed, it is impossible in normal surveys to have samples with the same incidence angle for various ranges and vice versa. This makes it also difficult to test the change trends by one of influencing factors (range or incidence angle) independently. Thus, it is assumed that all surfaces have Lambertian surface characteristics, which means that the influence by incidence angle can be handled with cosine law. Even if the assumption is not exactly met in natural surfaces, the incidence angle is not a negligible factor. As it was proven by the experiment that the decrease of intensity is significant when the incidence angle increases from 20 ° to 70 ° (Kukko et al., 2008), it can be considered that this assumption can deal with the effect on incidence angle. Secondly, in order to remove the bias by surface brightness, intensities in samples are divided by the relative surface reflectivity that is estimated by approximate surface brightness. The way to estimate relative reflectivity is same to pervious subsection (section 3.5.2).

After correcting the intensities of samples by cosine law and relative reflectivity, scatter plots are drawn by range and processed intensity data. As the bias by reflectivity was removed, samples from different surfaces are also combined in one plot. Also, the scatter plot is drawn by a set of the median values of ranges and intensities within sample groups due to variation of intensity value and error propagation in estimation procedures.

Finally, it is tried to find the best fit line between range and intensity data. The models for fitting line are flexibly selected depending on the pattern derived from the scatter plots. According to the selected model, parameters are estimated by available estimation methods such as the robust least square sum. By applying the same relative reflectivity to the same surface type in both sensors, intensities in both sensors are adjusted to each other. After the final model is defined with the estimated parameters, the entire dataset can be corrected.

### **3.6. Summary**

This chapter explained the methods based on two categories: theoretical model based approach and empirical model based approach. Three methods were derived in order to find the proper correction model for MLS data. In this study, it is a crucial part to inspect how the radar range equation works and to find the changing pattern of intensities in different ranges, incidence angles and sensors. However, the potential limitation exists in the unavoidable error propagation in each estimation procedure and the absence of a homogeneous area which can cover all extent of range and incidence angle that might occur in the entire point clouds. The whole procedure follows

Figure 3-1: Workflow of proposed method

and implementation and correction results are presented and discussed in next chapter.



## 4. Implementation and results

This chapter presents the implementation and results of the methods described in previous chapter. It consists of six sections. In section 4.1 the overview of dataset and study area are described with the optical property of the employed sensor system. Sections 4.2 ~ 4.5 are dedicated to explain how each step is implemented and show results by the proposed methods. Finally, the derived results are discussed with the concluding remarks (section 4.6).

### 4.1. Description of dataset

#### 4.1.1. Dataset and study area

The dataset for this research was collected by the LYNX Mobile Mapper (LYNX) made by the company Optech. The data provided consists of four sub-datasets which are point clouds (x, y, z coordinates, intensity, pulse number and time stamp), GPS data, offset data, and optical RGB images taken by mounted cameras. The data were surveyed in one of German villages in autumn of the year 2009. The detailed data related on circumstance such as temperature, visibility, humid, and pressures were not available. However, with the help of taken images it could be known that the weather was clear and surfaces were in dry condition during the data collection.

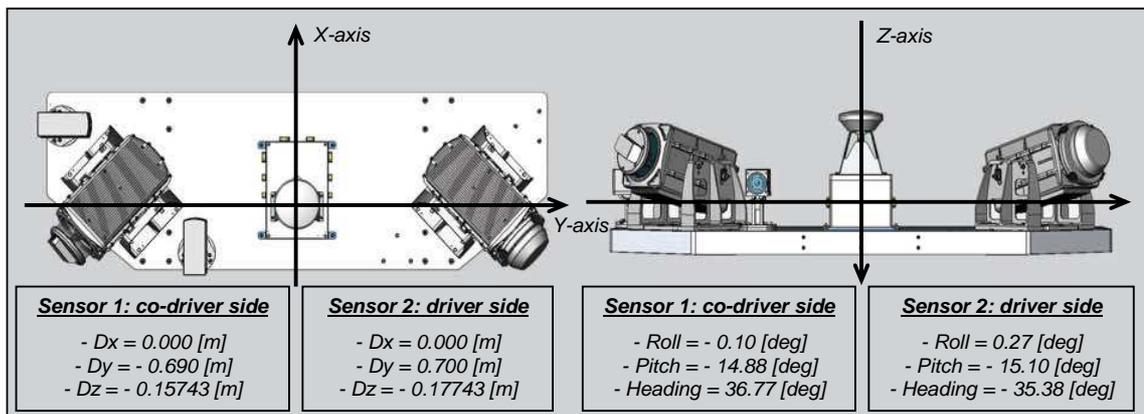


Figure 4-1: Offset of LYNX Mobile Mapper (Background image: (Optech, 2007))

As depicted in Figure 4-1, one or more (in this case two) laser scanners, two cameras, and GPS are mounted on LYNX platform. The laser scanners operate by the laser shot with the wavelength of 1,500 nm and the beam divergence of 0.23 mrad. Also, it is specified that a sensor can collect point data reflected from more than 100 m to 20% reflectivity target (Optech, 2007). The beam width at the beam exit of the transmitter, aperture diameter of the receiver and signal power at the emitted moment were not disclosed as usual data. The taken point clouds have the absolute accuracy of up to  $\pm 5$  cm depending on the accuracy of GPS data. The intensities are recorded as integer numbers by the lookup table of each sensor.

Figure 4-1 shows the coordinate system defined as a right-handed coordinate system where x-axis points into the driving direction, y-axis to the co-driver side, and z-axis downwards. Here, the offset data indicate the distance from GPS data to the sensor projection centre and the rotated angle of sensors in sensor coordinate system. Exactly, GPS data point the centre of the INS unit mounted in the middle of LYNX platform (the origin of sensor coordinate system in base coordinate system). While a vehicle is moving, GPS data are recorded with the frequency of 200 Hz. In addition, the offset values presented in Figure 4-1 are one that was used in this dataset. As the full sensor system is assembled whenever it is used, it can be different in every survey.

The dataset contains only one scan strip along a paved road where two sensors were used simultaneously. The estimated approximate total surveying time and distance were 66 second and 0.42 km respectively. The average and maximum speed of a vehicle were 23 km/h and 36 km/h and a vehicle stopped once for a while at a junction. The number of points by each sensor was similar and approximately 4 million. In order to test the correction models, two test data were selected as it is difficult to process large number of points at one goal. In Figure 4-2 and Figure 4-3, the test 1 and 2 are visualized. As the test data 1 includes several planer surfaces with homogeneous brightness such as a paved road, road marks, building walls painted in the same colour and building roofs, it was selected to compare the behaviour of intensities in different surface types. The test data 2 is chosen to inspect the behaviour of intensities of various ranges and incidence angles in a homogeneous surface. Each point cloud contains approximately 450,000 points. Sensor 1 and 2 mean driver side and co-driver side sensor respectively.

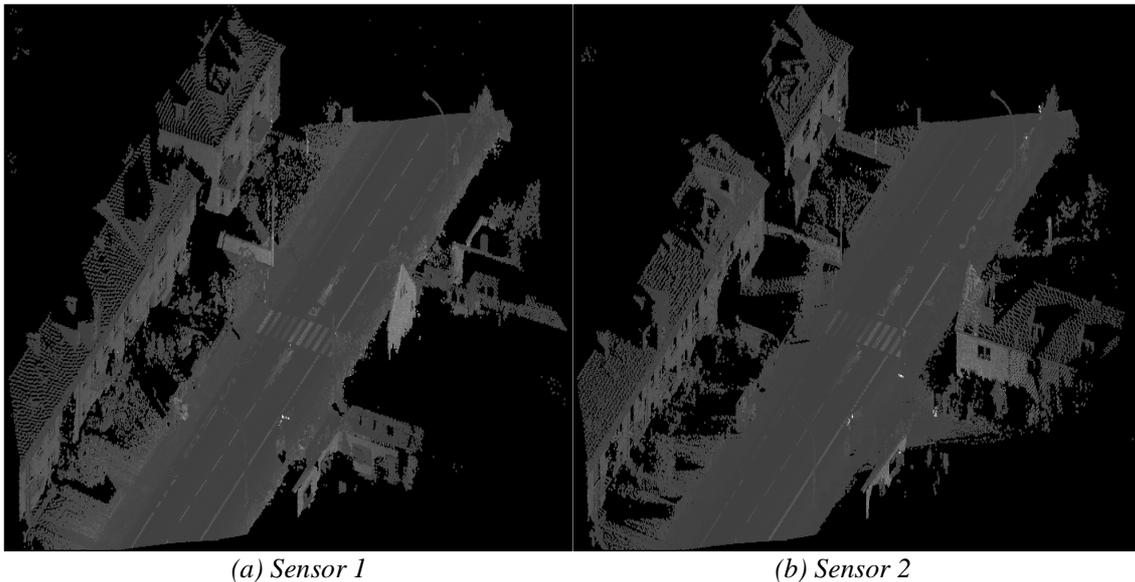
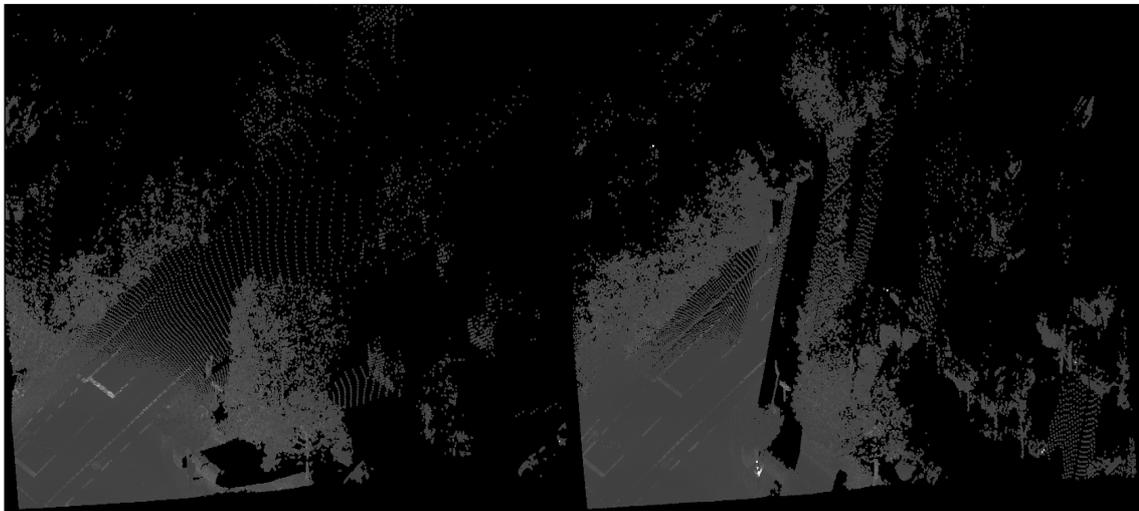


Figure 4-2: Test data 1 – road corridor with several planer surfaces with homogeneous brightness



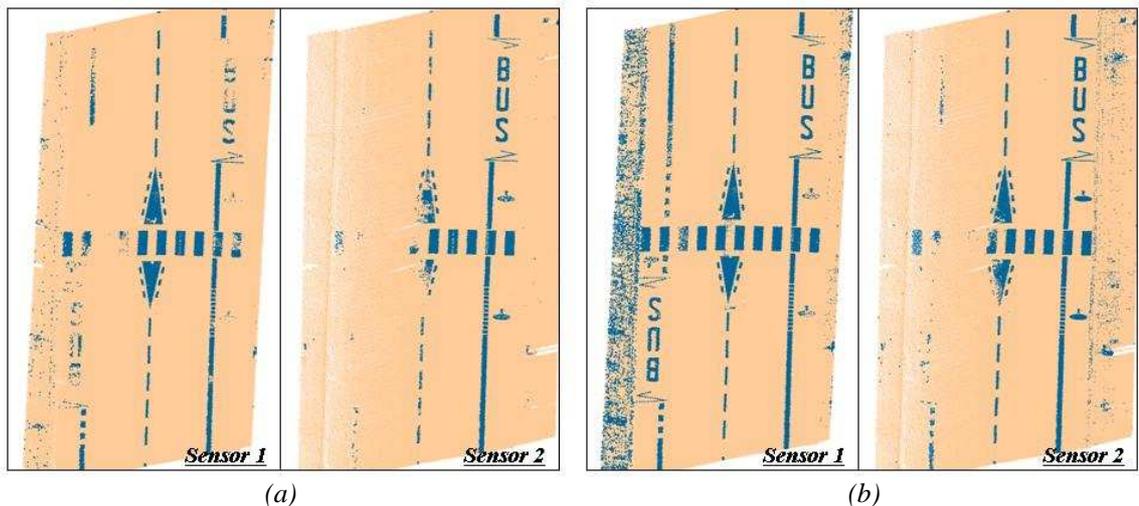
(a) Sensor 1

(b) Sensor 2

Figure 4-3: Test data 2 – Paved road surface with points of various ranges and incidence angles

#### 4.1.2. Examination of test data

In remote sensing, the reflectance from surfaces has been broadly used in the applications such as the classification of surface types and the image matching. According to this principle, MLS point cloud data can be used for the classification of the road surface. In this examination, test data 1 is used and it is anticipated that road marks can be retrieved easily by intensities since the contrast between paved road and road marks are quite clear in reality. In order to reduce the disturbance of surrounding points, road marks are classified from approximate ground points cropped by a certain height.



(a)

(b)

Figure 4-4: Road and road marks classified by different thresholds (a) 250~1000 and (b) 150~1000

In the classified results shown in Figure 4-4, it is observed that some of road marks are missing in (a) by smaller extent of thresholds. In (b) it is found that most of road marks are classified, however, some pedestrian road outside of the road is also classified into the road marks class despite of the different surface reflectivity. Moreover, it is observed that thresholds for data from sensor 1 are not

proper in data from sensor 2. Based on the test results, it can be considered that the cause is due to weakened intensities by increasing range and incidence angle. It is obvious that this would be more serious in further range. Also, it is indirectly known that measurements from sensors are different so that the adjustment between both sensors would be required.

## 4.2. Pre-processing

### 4.2.1. Implementation

As discussed in section 3.2, pre-processing is performed by filtering of multi-echoes and outliers (specular reflection). The filtering of multi-echoes is implemented with the help of pulse number (the count of pulses received from one emitted pulse) and last pulse flag (binary data to indicate the final echo collected from one transmitted pulse). The filtering of outliers is implemented by the histogram of intensities. In Figure 4-5, it is shown that most points are distributed in lower values and only few points have higher values. Therefore, its threshold is assigned as the sufficiently high value (upper 95% in a histogram), in order to minimize loss of data.

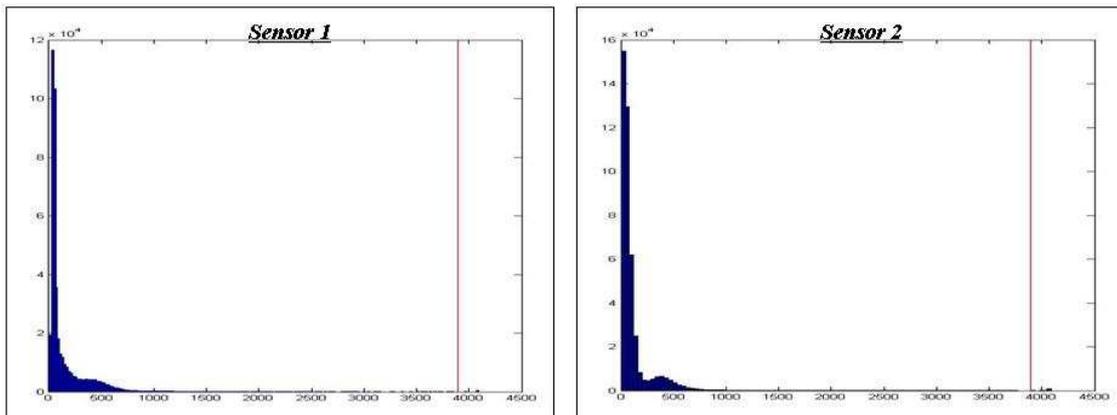


Figure 4-5: Histogram by intensities of test data 1 (red line: threshold, 95%)

### 4.2.2. Results

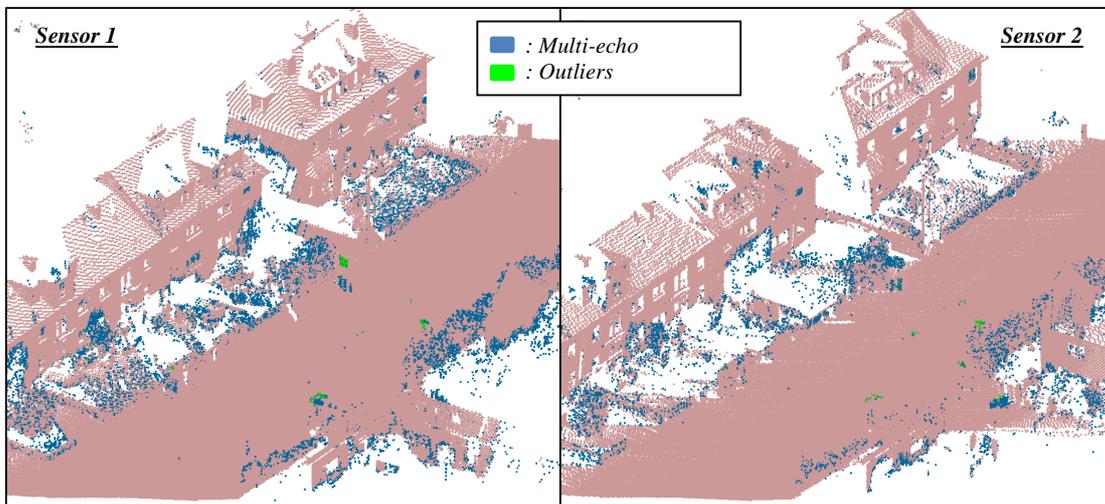


Figure 4-6: Filtering results of test data 1 (color-coded by filtering scheme)

<i>Sensor</i>	<i>Multi-echoes</i>	<i>Outliers</i>	<i>Remains</i>	<i>Total</i>
<i>1 (driver side)</i>	<i>18,587 (4.3%)</i>	<i>305 (0.1%)</i>	<i>417,891 (95.6%)</i>	<i>436,783</i>
<i>2 (co-driver side)</i>	<i>15,415 (3.5%)</i>	<i>636 (0.2%)</i>	<i>418,182 (96.3%)</i>	<i>434,233</i>

*Table 4-1: The number and proportion of filtered points in test data 1*

In Figure 4-6, it is observed that many multi-echoes exist around vegetation and moving objects and outliers are mainly distributed around flat metal surfaces. Table 4-1 indicates approximate 4.0 % of point clouds are filtered from test data 1 and it depends on the situation of surveyed area.

### 4.3. Reconstruction of scan geometry

#### 4.3.1. Sensor trajectory

As the correction of data is based on range and incidence angle, the estimation of the sensor trajectory plays an important role to reduce errors for fundamental data of the algorithm. This process is implemented by three steps: a) estimation of the coordinate system of sensor platform, b) rotation of offset data, and c) subtraction of rotated offset from GPS data.

Firstly, the coordinate system of sensor platform has to be derived which is defined x-axis as driving direction y-axis and z-axis as a right hand coordinate system (section 4.1.1). In order to estimate the driving direction of each GPS point, 50 points were selected by time stamp (previous 25 and next 25 points). The coordinate system can be defined from them by the robust least square sum. As the intervals of points are small, the estimation was properly operated in most places. However, incorrect results are produced while a vehicle is stationary to bend to a junction or wait for traffic signal (Figure 4-7 (a)). In this location the significant estimation errors result from insufficient points since most of the chosen 50 points by the time stamp are in the same position. Therefore, the points captured while stopping have to be filtered in order to reduce this kind of estimation errors. For implementation the least distance between the previous and next points is considered that can be regarded to be moving. In other words, if the distance is less than a certain threshold, the point is defined as stationary point and removed in further processes. Then based on the estimated x-axis, the coordinate system is formed by the right-hand rule from which the offset data is rotated by 3D rotation matrixes into the base coordinate system. Finally, the sensor position of each GPS point is calculated by subtracting rotated offset distance from the GPS point.



*Figure 4-7: Estimated sensor trajectory (a) without filtering stopping points and (b) after filtering*

In this study, the threshold is assigned as 1.5 cm, which is defined by trial-and-error since GPS error is different depending on the circumstance such as surveyed area and weather. In Figure 4-7, the bright white lines are the estimated trajectories of both sensors and the gray line between white lines is the GPS trajectory. Here, it is noticed that the estimation error of driving direction is shown as noises in the derived sensor trajectory and filtering can cause the break of trajectory. Therefore, the threshold has to be defined as the small value not to make the significant estimation error. The generated break line is extracted by the liner interpolation between two edge points.

#### 4.3.2. Surface normal

As discussed in section 3.3.2, surface normal is estimated by  $k$  nearest neighbours using KNN algorithm. For fast computation, the nearest neighbours are searched by ANN (approximate nearest neighbour) library written in the C++ programming language (Mount, 2006). The parameters required in ANN are importantly considered that include searching boundary and neighbourhood size in order to reduce the estimation error. Some factors have to be additionally taken into account such as point density, curvature of surface, noise (Mitra and Nguyen, 2003). However, since those kinds of factors in MLS are not uniform in overall region but flexible depending on range, vehicle speed and sensor, parameters are defined by trial-and-error within reasonable scope. In this research, the parameters are determined as spherical searching boundary of 0.5 m radius and 50 neighbourhood size considering high point density and relatively smooth surface curvature condition of the surveyed area. After searching 50 nearest neighbours, the surface normal is estimated by the planer regression. Two available equations discussed in section 3.3.2 are employed. During this process, some of points are filtered which cannot define surface normal due to the absence or lack of neighbours.

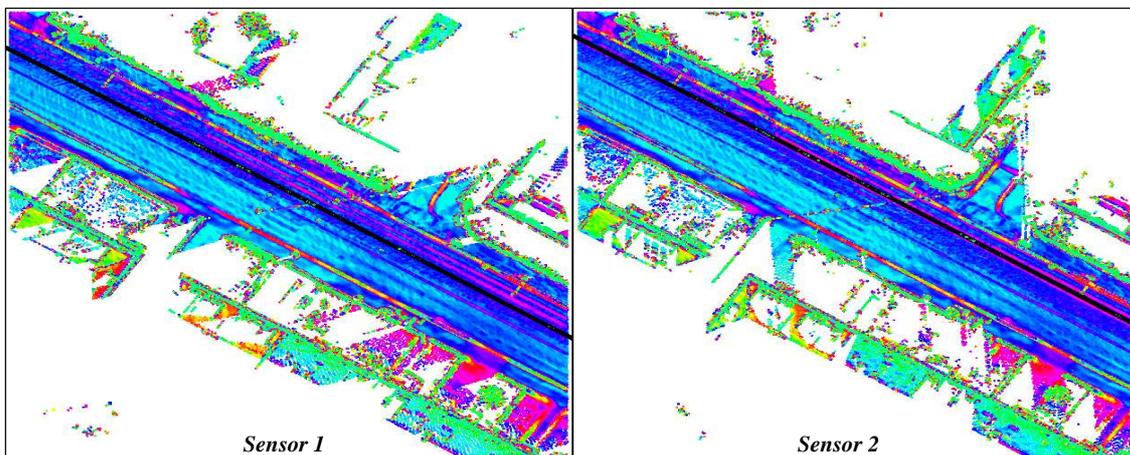


Figure 4-8: Estimated surface normal (color-coded by the angle with vertical axis)

In Figure 4-8, it is shown that points on planer surfaces have the similar values within the surface. Also observed is some noise around the sensor trajectory. It is considered that the noise along the sensor trajectory is caused from extremely high point density of surfaces with small ranges. Indeed, the reason of the noise is that the area by neighbours is smaller than the positional error of the point cloud. Thus, it can be thought of that the parameter on neighbourhood size needs to be differently applied according to the local point density of point clouds. However, this will be left as further study since the estimation error of surface normal in noisy point cloud data cannot be completely removed even after adjusting parameters (Mitra and Nguyen, 2003).

### 4.3.3. Range

The calculation of ranges is implemented by two steps: a) the definition of a laser shot vector from target to sensor and b) the calculation of length of the laser shot vector. According to these steps, the laser shot vector is defined by matching of each target point and the corresponding sensor point by using the time stamp which is included in point cloud data and GPS data. The exact sensor point is derived by linear interpolation of two points with the closest time stamp. Finally, the range of each point is calculated by the length of the laser shot vector. The points from a range of greater than 100 m are filtered. (100 m is the maximum range of LYNX stated in the sensor specification.)

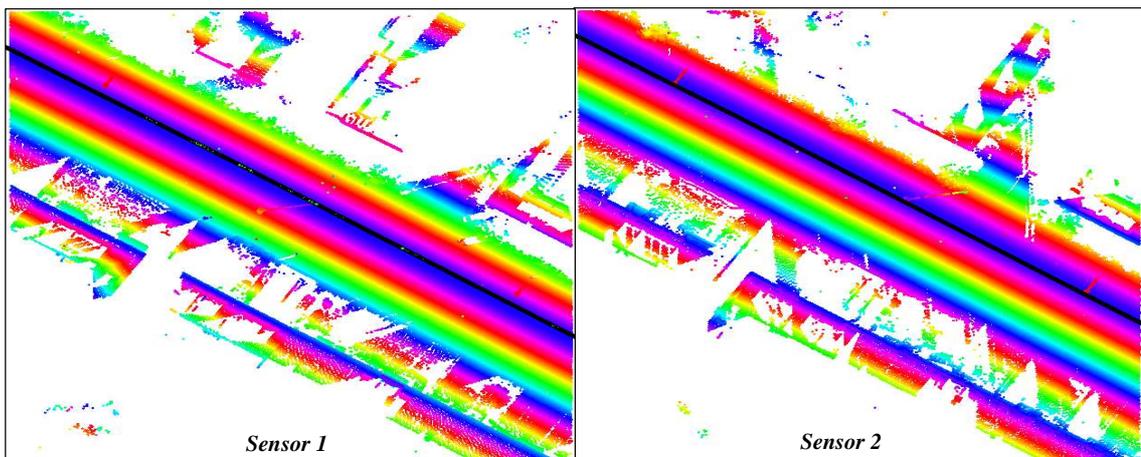


Figure 4-9: Estimated ranges (color-coded by ranges, blue line: sensor trajectory)

Figure 4-9 shows the colour coded point cloud by increasing range, which is slightly shifted from the sensor trajectory. It can be considered that the shifted colour fringe results from the rotated angle of sensors specified in the offset data. Additionally, there is no point with a range of more than 100 m.

### 4.3.4. Incidence angle

Up till now, the required data for incidence angle had already been processed in previous steps. In this step, the angle between two vectors (laser shot vector and surface normal) is calculated. As the angle can be negative or greater than  $90^\circ$  due to the direction of vectors, these values are converted into a range between  $0^\circ$  and  $90^\circ$ .

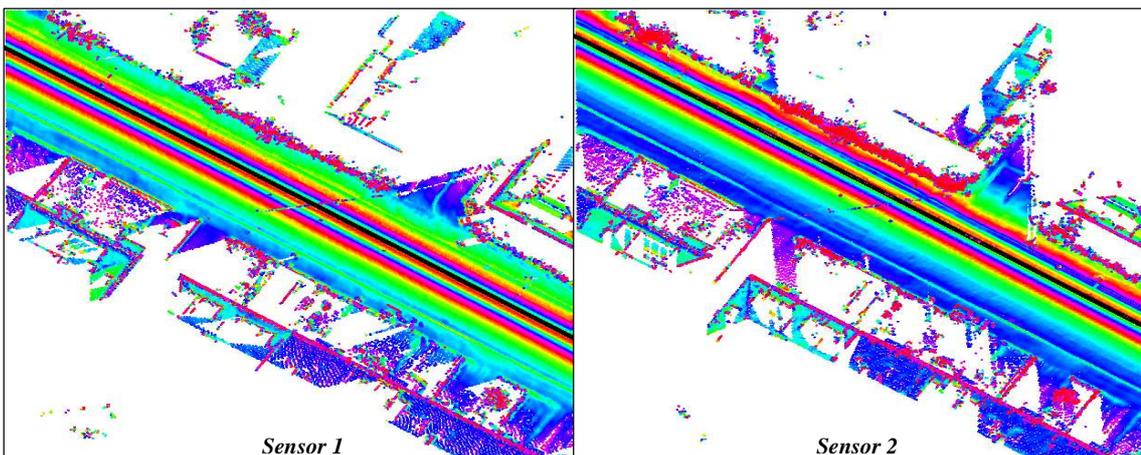


Figure 4-10: Estimated incidence angles (color-coded by incidence angles)

In Figure 4-10 the similar colour fringe to Figure 4-9 is observed which is parallel to the sensor trajectory. The different patterns are found in extruding objects such as buildings and vegetation. In addition, it is observed that the noises from the estimation of surface normal are also kept in incidence angle.

#### 4.4. Data correction based on radar range equation

This step is implemented by replacing the estimated range and incidence angle in equation (2.9),  $\rho \propto R^2 \cdot 10^{2Ra/10000} \cdot \cos^{-1} \alpha \cdot C$ . Based on weather condition and Table 3-2, air attenuation coefficient ( $a$ ) is assigned as 0.2 dB/km. Constant factor ( $C$ ) is set up as  $10^2$  for the corrected results to have the appropriate scope of intensities for visualization.

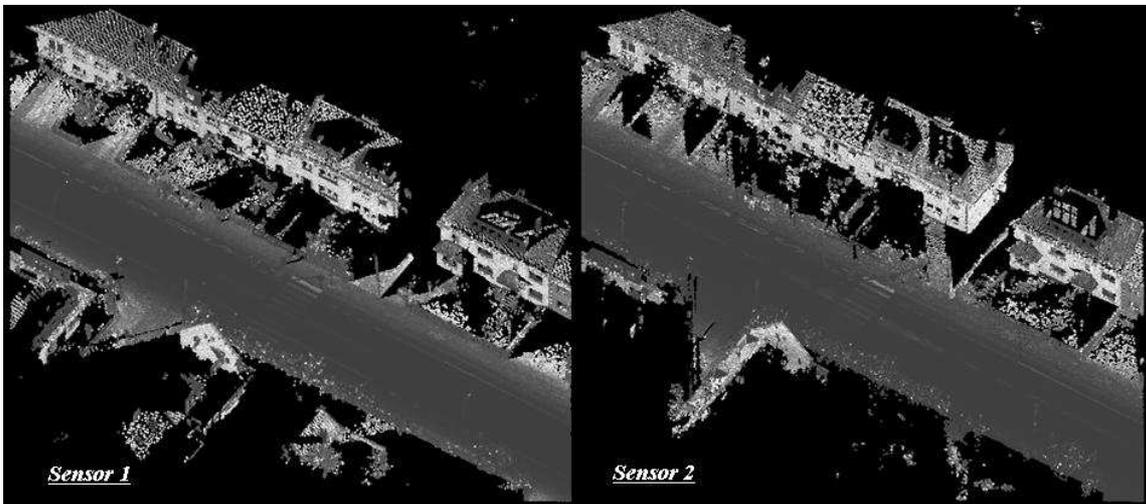


Figure 4-11: Correction results by theoretical model

Figure 4-11 shows the corrected results of two sensors. Compared with uncorrected data (Figure 4-2), the contrast between different surfaces becomes worse on road surfaces of near range around the sensor trajectory. It is also observed that intensities are much larger than the expected in building walls and roofs of far range. The problem of this method is that the corrected intensities on the surface with the same properties look different depending on their calculated ranges. For example, in zebra road marks for crossing, intensities in near range are less bright than those in far range. Moreover, some building roofs have higher values than road marks even though the brightness of roofs is much less than road marks in reality. Consequently, it can be concluded that intensities are over-estimated in far range by the simplified radar range equation. In addition, it is visually proven that the radar range equation cannot apply to MLS data as well as TLS data (Pfeifer et al., 2008).

The similar results are also found in test data 2 and the results up to this process are presented in Appendix A: Correction of test data 2 by the radar range equation. In chapter 5, the result will be re-evaluated in detail.



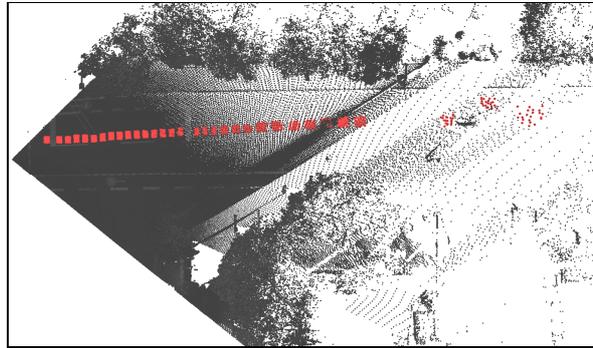


Figure 4-13: Sampling for defining empirical model (test data 2)

As shown in Figure 4-12 and Figure 4-13, five widely spread surfaces were picked and chosen in test data 1 and only paved road surface with asphalt was sampled in test data 2.

### Analysis of samples

After sampling, the change of intensities with respect to range and incidence angle is separately inspected by each sensor. Basically, this analysis is done by box plots due to the variation of intensity data. Since sample sets have been selected from the point cloud combined with data from both sensors, firstly a sample set is divided into each sensor. Then, the sample groups with less than 10 points are removed in order to increase the reliability of sampled dataset. Finally, box plots are drawn by sample groups arranged by median ranges in order to observe the pattern of intensities according to increasing range.

Sensor	Surface	Box plot	Distribution of median values		
			Range	Incidence	Intensity
1	Asphalt		2.4 ~ 7.4	18.8 ~ 73.2	19 ~ 112
	Road marks		2.5 ~ 7.4	22.3 ~ 72.0	161 ~ 695
	Pedestrian		5.7 ~ 10.5	63.0 ~ 76.5	77 ~ 228

Sensor	Surface	Box plot			Distribution of median values		
					Range	Incidence	Intensity
1	Wall				14.0 ~ 14.9 20.7 ~ 22.1	34.0 ~ 39.0	433 ~ 847 1068 ~ 1169
	Roof				14.3 ~ 18.6 22.5 ~ 29.0	60.9 ~ 64.8 70.6 ~ 79.7	24 ~ 357
2	Asphalt				2.5 ~ 10.2	20.0 ~ 78.2	5 ~ 76
	Road marks				2.5 ~ 9.0	18.8 ~ 75.9	62 ~ 634
	Pedestrian				3.3 ~ 5.2 11.1 ~ 13.2	43.2 ~ 61.1 75.1 ~ 79.0	73 ~ 110
	Wall				12.1 ~ 12.2 13.9 ~ 15.3 22.3 ~ 23.0	34.9 ~ 54.1	344 ~ 902
	Roof				12.9 ~ 18.8 22.6 ~ 27.2	61.4 ~ 78.5	28 ~ 256

Table 4-2: Box plots by sample sets (data set 1)

Sensor	Surface	Box plot			Distribution of median values		
					Range	Incidence	Intensity
1	Asphalt				2.4 ~ 17.0	20.7 ~ 82.3	22 ~ 68
2	Asphalt				2.4 ~ 24.1	18.4 ~ 85.2	3 ~ 50

Table 4-3: Box plots by sample sets (data set 2)

Table 4-2 and Table 4-3 show the box plots of sample sets by range, incidence angle and intensity. The following trends can be retrieved:

- As expected, it is found in road-related sample sets (asphalt, road marks and pedestrian road) that the incidence angle is simultaneously increased while the range is increasing. Thus, there is no sample group that have large incidence angles in the nearest range and vice versa. Also, in the range plots of building-related sample sets (walls and roofs) it is observed that sample groups are divided into two parts depending on the place of buildings (left or right side of the road). The incidence angles are almost constant due to vertical slopes of walls and comparably leaned slopes of roofs. It results from the characteristics of surveyed area by MLS that collects road-corridors from close distances. Consequently, the limitation in sampling is reconfirmed that whole extent of independent variables (range and incidence angle) cannot be inspected in normal point clouds.
- In intensity plots, the fluctuating patterns of intensities are found in both sensors. Namely, in sensor 1 intensities are decreased until 5 m, increased in 5 ~ 8 m and decreased again from 8 m. Also, in sensor 2 they are decreased until 8 m, increased in 8 ~ 15 m and decreased again from 15 m. The fluctuating degrees are different according to surface types, however, the comparable patterns are observed in every plot.

#### 4.5.2. Adjusted radar range equation

As mentioned in section 3.5.2, equation (3.1),  $\tilde{\rho} = I \cdot R^a \cdot 10^{2bR} \cdot \cos^c(\theta) \cdot 10^d$  is employed which includes four parameters. In the equation the relative reflectivity ( $\tilde{\rho}$ ) is determined as the approximate value since the surveyed area does not include the reference surface and the exact quantification of natural surfaces is not possible in this case. For this, sample groups with similar range and incidence angle are selected from each surface types based on the drawn box plots in Table 4-2 and Table 4-3. It is determined by median of intensities of the chosen sample groups. Moreover, in order to adjust the difference of measurements of each sensor, relative reflectivity only from a sensor is identically used to both sensors. Therefore, it can be expected that corrected results of both sensor

data would have the similar values and the additional process for adjustment is not necessary. Using the logarithmic least square sum, the four parameters are estimated with a representative set (medians) of ranges, incidence angles and intensities within sample groups.

Data set	Sensor	Surface	$\hat{\rho}$	$a$	$b$	$C$	$d$
1	1	Asphalt	40	3.51	-0.198	-0.70	-0.60
		Road marks	380	1.68	-0.223	-3.07	0.13
		Pedestrian	100	-8.62	0.214	-0.03	4.15
		Wall	450	-4.18	0.075	4.48	2.73
		Roof	250	8.36	-0.057	-0.04	-8.21
	2	Asphalt	40	1.00	-0.006	-0.50	-0.51
		Road marks	380	8.29	0.008	6.77	-3.37
		Pedestrian	100	1.40	-0.031	-0.25	-0.60
		Wall	450	0.81	0.003	0.36	-1.18
		Roof	250	1.93	0.004	0.02	-2.21
2	1	Asphalt	40	-0.36	0.011	-0.18	0.02
	2	Asphalt	40	-20.48	-0.028	-21.11	7.70

Table 4-4: Estimated parameters according to sample sets

In Table 4-4 it is shown that the estimated parameters can be different depending on the sensor. Also, the totally different estimation parameters are shown even in asphalt surface and road marks with similar extent of range and incidence angle. The different distribution of range and incidence angle in sample sets can be considered as the possible reason. Moreover, it is noted that the mentioned problem on extrapolation can be more serious in this method. Hence, it is proposed that the final model needs to be selected considering the distribution of range and incidence angle in a sample set.

Based on the distribution of range and incidence angle within each sample set, the final equations are selected.

Data set	Sensor	Applied range (m)	Final model
1	1	0 ~ 7.4	$\rho \propto R^{3.51} \cdot 10^{2R \times (-0.198)} \cdot \cos^{-0.70} \alpha \cdot 10^{-0.60}$
		7.4 ~ 12.3	$\rho \propto R^{-8.62} \cdot 10^{2R \times (0.214)} \cdot \cos^{-0.03} \alpha \cdot 10^{4.15}$
		12.3 ~ 100	$\rho \propto R^{8.36} \cdot 10^{2R \times (-0.057)} \cdot \cos^{-0.04} \alpha \cdot 10^{-8.21}$
	2	0 ~ 10.2	$\rho \propto R^{1.00} \cdot 10^{2R \times (-0.006)} \cdot \cos^{-0.50} \alpha \cdot 10^{-0.51}$
		10.5 ~ 12.9	$\rho \propto R^{1.40} \cdot 10^{2R \times (-0.031)} \cdot \cos^{-0.25} \alpha \cdot 10^{-0.60}$
		12.9 ~ 100	$\rho \propto R^{1.93} \cdot 10^{2R \times (0.004)} \cdot \cos^{0.02} \alpha \cdot 10^{-2.21}$
2	1	0 ~ 100	$\rho \propto R^{-0.36} \cdot 10^{2R \times (0.011)} \cdot \cos^{-0.18} \alpha \cdot 10^{0.02}$
	2	0 ~ 100	$\rho \propto R^{-20.48} \cdot 10^{2R \times (-0.028)} \cdot \cos^{-21.11} \alpha \cdot 10^{7.70}$

Table 4-5: Applied equations according to range

Table 4-5 presents the defined final equations. In inspecting the distribution of medians in Table 4-2, two sample sets were included in near range (0 ~ 7.4 m) of sensor 1. As their distributions of range

and incidence angle are compared, it is observed that asphalt sample set are more stably distributed over its extent whereas road marks sample set is missing some parts. Therefore, the parameters for 0 ~ 7.4 m are decided by the asphalt sample set. As only pedestrian road surface exists in middle range (7.4 ~ 10.5 m), the parameters for 7.4 ~ 10.5 m are defined by the pedestrian road surface. Finally, the parameters for far range (14.0 m ~) is defined by the building roof surface since the roof sample set is more widely distributed than the building wall sample set. In the same way, the model of sensor 2 is also determined. As test data 2 has a sample set that is distributed in various ranges and incidence angles, it is defined by one equation for each sensor.

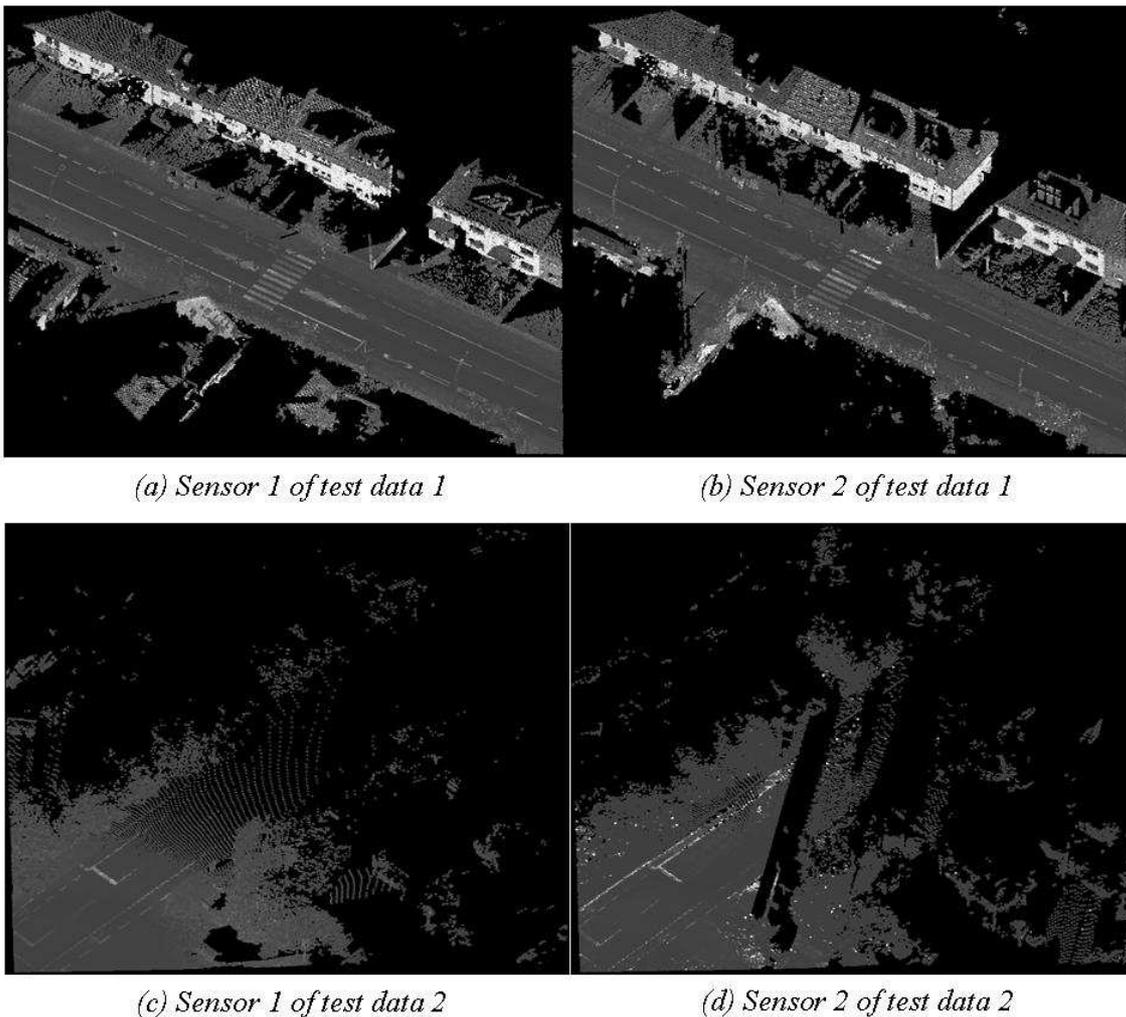


Figure 4-14: Correction results by adjusted radar range equation

Figure 4-14 shows the corrected results of test data 1 and 2. Visually compared to uncorrected data (Figure 4-2, Figure 4-3) and the result by radar range equation (Figure 4-11), it is observed that the contrast between different surfaces becomes more significant. However, the performance in far range needs to be more inspected as bright speckles are detected in some road marks of Figure 4-14 (d). More evaluation is required and it will be done in chapter 5.

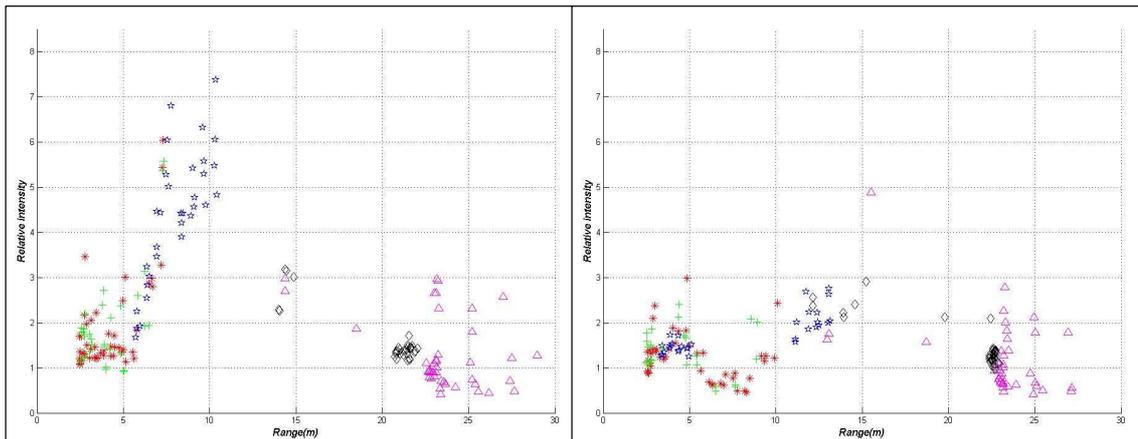
### 4.5.3. Data fitting

In this procedure, the same sample sets are applied and the behaviour of intensities is separately analyzed with each sensor as well. As the first step, samples are processed by Lambertian cosine law and surface brightness in order to remove the bias of samples caused from incidence angle and surface reflectivity.

$$I_P = \frac{I}{\cos(\theta) \cdot \hat{\rho}} \quad (4.1)$$

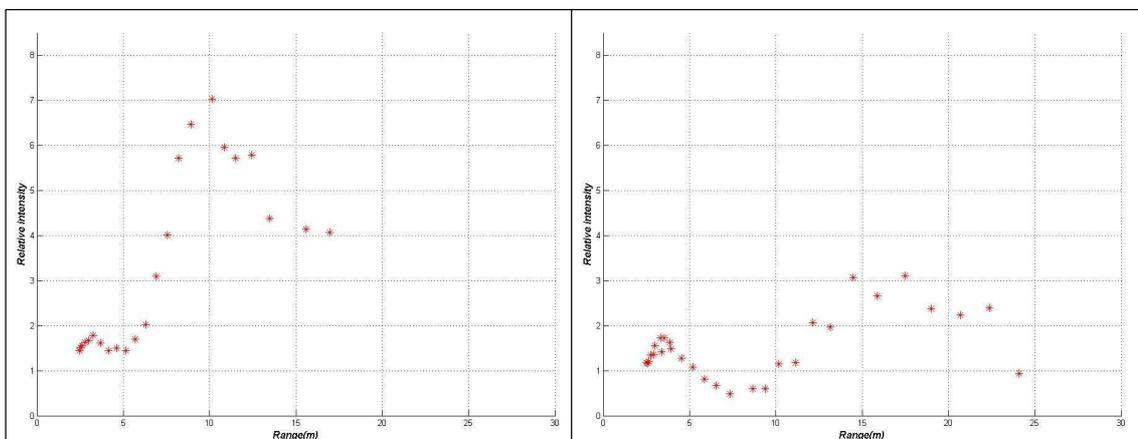
Where,  $\hat{\rho}$  = relative reflectivity of sample set

The relative reflectivity is estimated from data set itself in the same way to previous method. Thus,  $\hat{\rho}$  in Table 4-4 is applied to each sample set. After removing the bias by equation (4.1), the processed samples are drawn in one scatter plot as the relation between range and intensity. In the scatter plot points are drawn by representative values (medians) of each variable within sample groups.



(a) Sensor 1 of test data 1

(b) Sensor 2 of test data 1



(c) Sensor 1 of test data 2

(d) Sensor 2 of test data 2

Figure 4-15: Scatter plot by range and processed relative intensity (red \*: asphalt, green+: road marks, blue \*: pedestrian road, black ◊: building wall and magenta △: building roof)

Figure 4-15 is divided into four plots by sensor and test data. From the plots it is proven that the behaviours of both sensors are clearly distinguished. Each sensor does not show the significantly different behaviour in different regions. On the pattern of intensity with respect to range, the change of both shows decrease after slight increase in near range and intensity is stably decreased in far range. However, the behaviour of intensities in middle ranges between approximately 5 ~ 15 m is different according to each sensor. It is shown that they act even into the opposite side, or decrease after steep increase in sensor 1 and increase after slow decrease in sensor 2. Hence, it has to be suggested that the correction models are separately applied to each sensor. They can be used to the entire data as far as the data was collected in the same survey.

Sensor	Range (m)	Equation
1	0 ~ 6.0	$I_c = 0.13 \cdot R^3 - 1.59 \cdot R^2 + 6.17 \cdot R - 6.03$
	6.0 ~ 10.7	$I_c = 1.27 \cdot R - 5.58$
	10.7 ~	$I_c = \frac{902.63}{R^2}$
2	0 ~ 8.7	$I_c = -0.01 \cdot R^4 + 0.21 \cdot R^3 - 1.97 \cdot R^2 + 7.53 \cdot R - 8.37$
	8.7 ~ 16.9	$I_c = 0.32 \cdot R - 2.22$
	16.9 ~	$I_c = \frac{924.48}{R^2}$

Table 4-6: Correction model based on data fitting

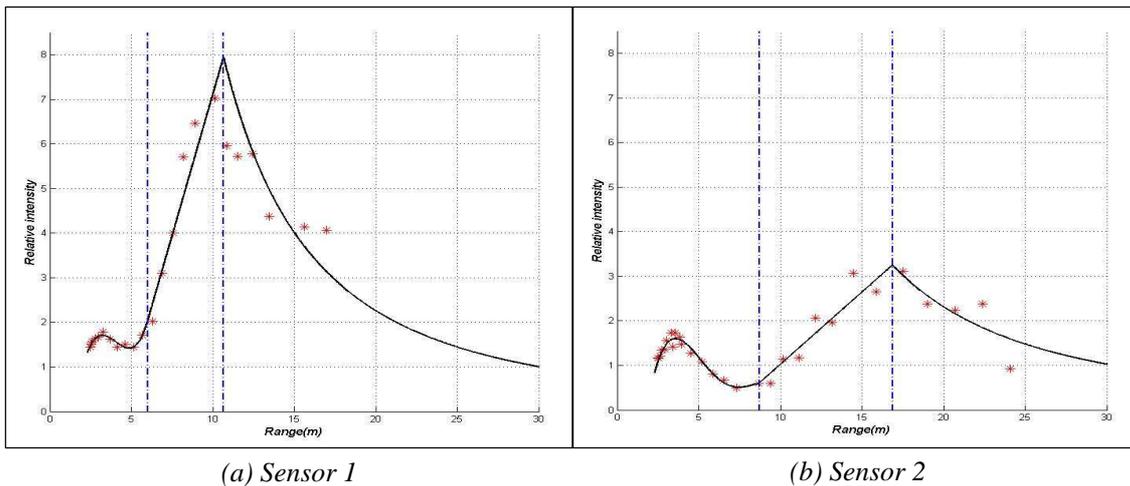


Figure 4-16: Fitting lines on both data

According to the pattern retrieved from plots, the correction model is determined as the third or fourth polynomial equation in near range, linear equation in middle range and the inversely second order polynomial equation in far range. As data from two regions have the similar pattern, both data can be simultaneously considered. Since test data 1 was, however, adjusted by approximate surface brightness and test data 2 still includes sufficient sample groups from the unique surface, the fitting line is estimated based on test data 2. The robust least square sum is employed in a parameter

estimation procedure. Table 4-6 describes the final correction model in this research. Also, Figure 4-16 presents the fitting lines on scatter plots by range and relative intensity of test data 2 used in modelling. It is shown that this plot has much difference with the smooth curve simulated by simplified radar range equation (Figure 3-5). However, this fitting line is still limitation especially in the second sharp peak due to the linear equation of middle range. Since natural phenomena usually do not have this kind of sharp change, corrected intensities around this part would have more errors. Therefore, it needs to be considered that this model is only one example to apply to this study. In order to reduce residuals of estimation, it must be tested by more sophisticate fitting techniques to express the found pattern.

Then, the correction of the entire dataset is implemented by dividing intensity of each point by cosine of its incidence angle and the calculated correction factor ( $I_C$ ) by correction model according to its range (Table 4-6).

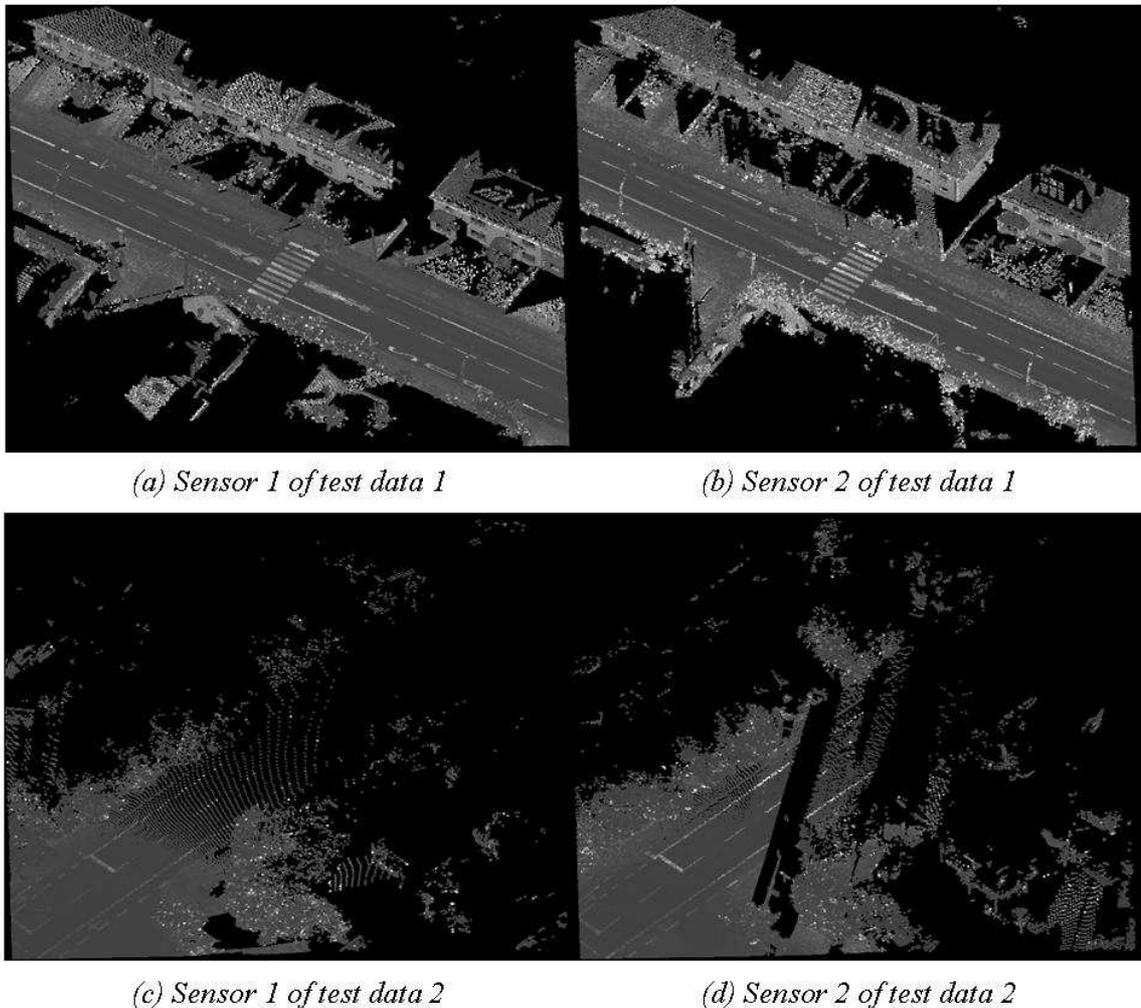


Figure 4-17: Correction result by data fitting

Figure 4-17 shows the corrected results of test data 1 and 2. As the model based on the adjusted radar range equation has shown, the results are more significant than the uncorrected data and the result by the radar range equation in the aspects of the contrast between different surfaces. However, some

speckles are also observed around the corners of planer surfaces and vegetation. More detailed assessment will be done in the following chapter 5.

#### **4.6. Concluding remarks**

In this chapter, the adopted methods were implemented and three sets of results were presented using two different regions within the same dataset. Based on visual inspection of uncorrected data and three results (Figure 4-2, Figure 4-11, Figure 4-14 and Figure 4-17), visual contrast of different surfaces could be investigated. It can be concluded that the radar range equation makes intensities in near range to be compressed and intensities in far range to be exaggerated. On the other hands, it is found that the adjusted radar range equation and the data fitting have more significant contrast between different surface types compared to the theoretical equation.

However, during processing errors in estimation procedures have been found. Especially, this variation is more seriously observed in surface normal estimation of regions with extremely high point density as well as around corners and edges of artificial objects. As this estimation error is propagated to correction results, the application of methods is restricted to planer surfaces. Furthermore, the tests by more sophisticate fitting techniques are essential in order to improve the proposed correction models.

## 5. Evaluation and analysis

In chapter 4, the correction results have been reproduced and a simple assessment was addressed by visual inspection. Thus, this chapter is dedicated for systematic evaluation of correction results, analysis of the possible causes for incorrect results and discussion on limitations of correction models. The performance evaluation of each model is carried out in two ways: a) statistical analysis by samples and b) analysis by comparison of classified surface types using corrected intensities. Thus, each following section describes statistical analysis by samples (section 5.1), analysis by classification of surface types (section 5.2) and concluding remarks (section 5.3).

### 5.1. Statistical analysis based on samples

#### 5.1.1. Evaluation scheme

As this evaluation method provides quantified result, it has been widely used in assessment. In this step another sampling is carefully carried out for the evaluation of correction models since samples in an evaluation procedure also play a crucial role just like that in modelling procedure. The sampling scheme is same as that in sampling step mentioned in section 3.5.1. That is, samples are chosen a) manually, b) by stratified sampling scheme, c) from regions with various ranges and incidence angles, and d) from smooth planer surfaces.

However, there is still some limitation in sampling due to the sampling error in manual sampling and estimation error. As noticed in section 4.3.2, the estimation of surface normal using nearest neighbours was not completely overcome in the nearest ranges with high point density. Also, it is not guaranteed that all sampled points are from the homogeneous surface with unique reflectivity. In order to minimize bias from the error unrelated on correction models, the outliers need to be excluded from sampling subjects. Hence, correction models are evaluated using intensities between 1<sup>st</sup> and upper 99<sup>th</sup> percentiles and median are more taken into account than average.

After careful sampling and filtering outliers, the intensities are analyzed by following four aspects:

- Consistency of intensities in the unique surface type: As a sample set consists of sample groups taken from homogeneous surface, it can be expected that those intensities have similar values after correction. This expectation is one of fundamental purposes to correct intensity data from which the model performance can be evaluated by the comparison of medians of sample groups. The assessment results of correction models are figured out by box plot and the standard deviation between medians of sample groups. Moreover, this aspect is essential since the correction model is meaningless if similarity of intensities in the same surface is considerably low. Thus, the priority is given to this aspect in overall evaluation.
- Consistency of both sensors: Both sensors were produced by the same manufacturer and take point clouds at the same time. However, the taken intensities show different pattern due to sensor own systematic factor. The processes have been independently modelled and corrected.

However, most laser scanning applications simultaneously use point clouds from both sensors and even from different strips. For further applications it is essential that intensities from both sensors have to be adjusted to each sensor. The consistency of both sensors is demonstrated by the distribution of intensities collected by two sensors from homogeneous surface.

- Contrast between different surface types: Since intensity data are relative values rather than absolute signal power, contrast between different surfaces is important for intensity-based applications. It is verified if intensities between similar surface types are similar and if the contrast between different surface types is clear. However, as mentioned before, studied dataset is from natural surface and does not include reference targets. Thus, the exact proportion of surface reflectivity cannot be quantified. Only the approximate surface brightness is applied which is observed from images by mounted cameras.
- Local noise reduction of intensities: Even if it is supposed that the effect of noisy characteristics of laser scanning data can be ignored by filtering process, uncorrected intensities still vary due to the influence by range and incidence angle. In previous evaluation, the variation in overall region was inspected whereas the variation in small local regions is investigated in this evaluation. Therefore, the evaluation of this aspect is presented by comparison of standard deviation of intensities within a sample group (patch).

## **5.1.2. Results analysis**

### **5.1.2.1. Sampling for evaluation**

According to mentioned sampling schema, sample sets are created. In test data 1, four widely spread surfaces are examined: paved road with asphalt, road marks for crossing, building wall painted in white and building roof. In test data 2, only asphalt surface is examined. In order to minimize the noise caused from mistakes in manual sampling, the images are referred which are taken from mounted cameras.

Figure 5-1 shows five selected sample sets (four from test data 1 and one from test data 2) from homogeneous surfaces. Each set consists of ten sample groups that are chosen using small rectangular patches. As a sample group represents model performance in a certain range and incidence angle, it is intended that a sample group has less diverse range and incidence angle. The median is used as the representative of range and incidence angle for a sample group. Table 5-1 shows the distribution of medians of ranges and incidence angles that each sample set has. The standard deviation is less than 0.4 m and 3° respectively for range and incidence angle within sample groups. In addition, it is intended that the (a) and (b) sample sets are distributed in the near range (below 10 m), (c) and (d) sets in the far range (above 10m), and (e) sample set in the various ranges in order to inspect the performance in a certain range.

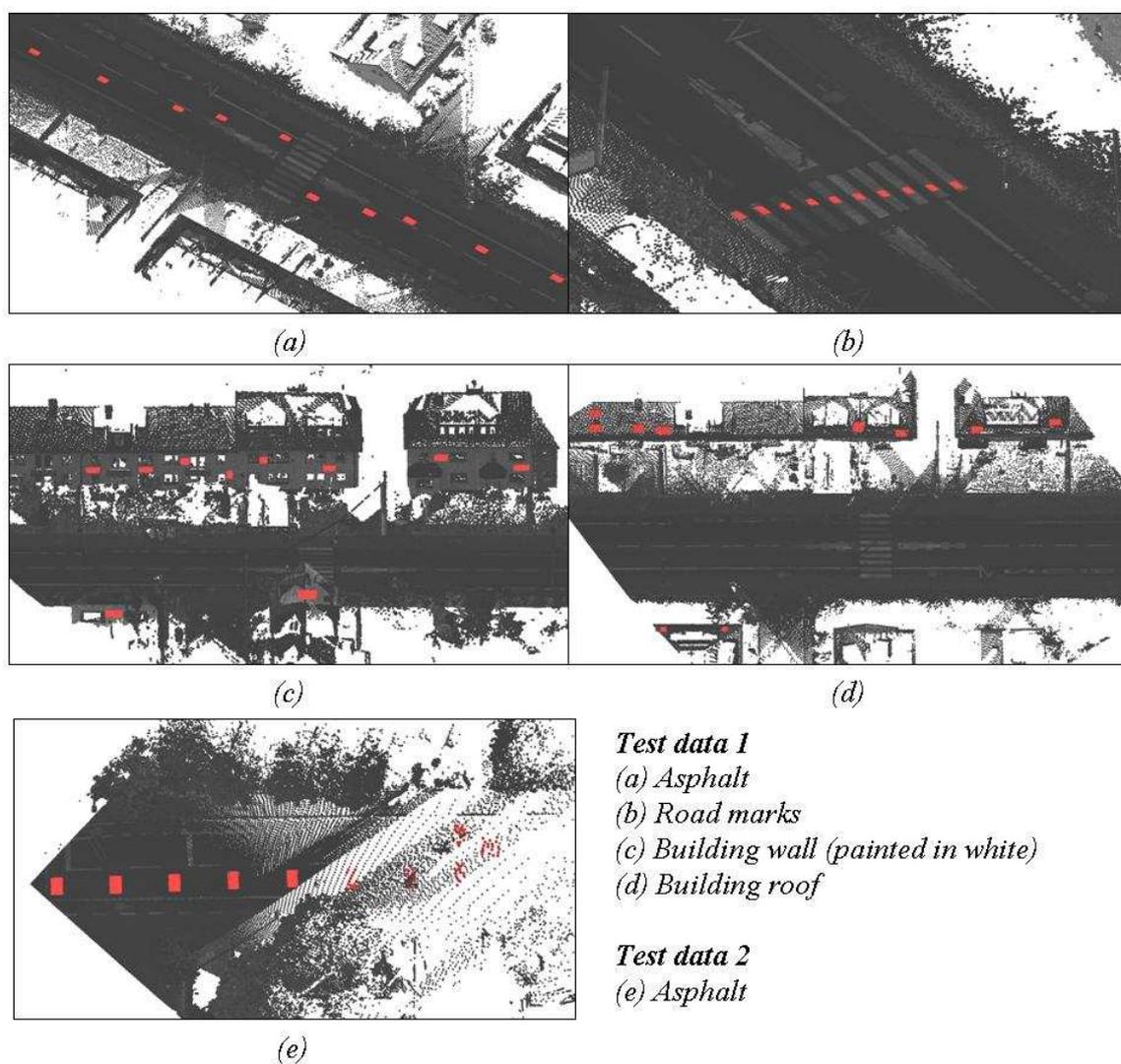


Figure 5-1: Sampling for evaluation

Sensor	Variable	(a)	(b)	(c)	(d)	(e)
1	Range (m)	2.4 ~ 5.2	2.5 ~ 7.4	14.0 ~ 22.1	14.3 ~ 26.5	2.6 ~ 38.4
	Incidence angle (°)	18.8 ~ 63.9	22.0 ~ 72.6	33.8 ~ 39.1	60.5 ~ 78.8	27.5 ~ 86.8
2	Range (m)	2.6 ~ 7.8	2.5 ~ 10.1	12.1 ~ 22.9	12.9 ~ 26.1	3.2 ~ 42.6
	Incidence angle (°)	21.9 ~ 73.1	19.0 ~ 77.4	35.2 ~ 31.5	61.5 ~ 77.8	42.1 ~ 87.1

Table 5-1: Distribution of medians of ranges and incidence angles

### 5.1.2.2. Evaluation results

#### Consistency of intensities in the unique surface type

In order to verify if sample groups in a unique surface are similar, intensities of each sample group are compared. In Figure 5-2, each figure consists of ten box plots which are drawn by ten sample groups in a homogeneous surface. The x-axis of figures is the sample groups arranged by median of ranges of

points within a sample group and the y-axis is relative intensity value. Thus, figures can be regarded as a plot of relative intensity by a function of range. However, it has to be kept in mind that the intervals of median range between sample groups are different. The values below plots are derived by standard deviation of ten medians that are calculated within each sample group. From the values, the general consistency between sample groups can be outlined by standard deviation in Figure 5-2.

Sensor	Surface	Uncorrected data	Radar range equation	Adjusted radar range equation	Data fitting
1	(a) Asphalt				
		SD: 13.6	SD: 3.5	SD: 10.7	SD: 7.0
	(b) Road mark				
		SD: 144.3	SD: 318.7	SD: 112.1	SD: 131.5
	(c) Wall				
	SD: 225.6	SD: 352.2	SD: 353.9	SD: 35.8	
(d) Roof					
	SD: 107.5	SD: 304.5	SD: 76.2	SD: 33.6	
(e) Asphalt					
	SD: 18.0	SD: 250.9	SD: 15.4	SD: 20.8	

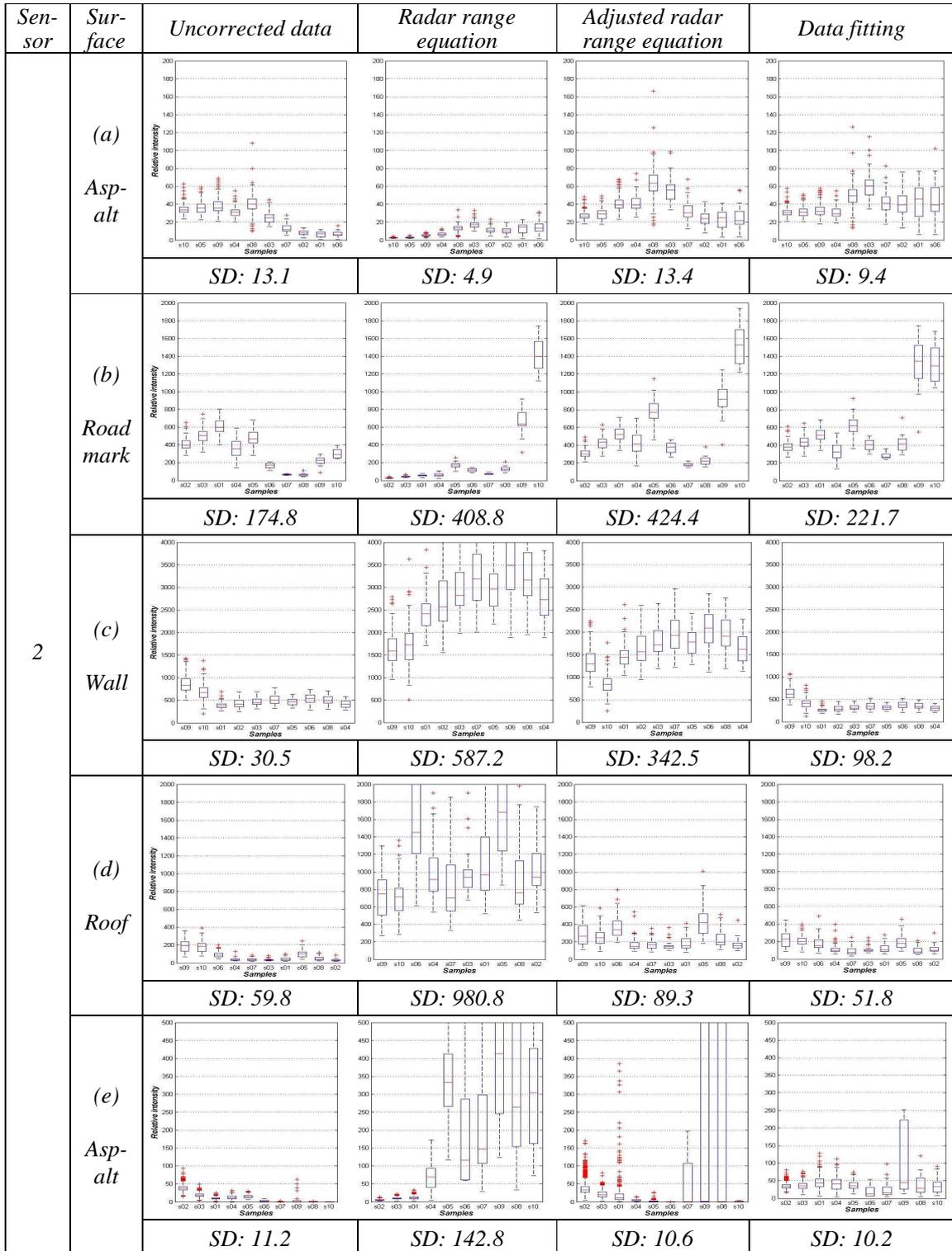


Figure 5-2: Comparison of sample groups in homogeneous area

For general consistency, it is observed that the radar range equation makes smaller the variation of (a) asphalt surface of the nearest range (below 5 m) whereas the variation of other sample groups is increased in further range (above 5 m). Moreover, it is shown that their plots are even fluctuating in far range. The adjusted radar range equation improves consistency in (a) asphalt, (d) roof and (e)

asphalt surfaces, but other surfaces have larger variation. On the other hands, in the data fitting it is shown that variance of most surfaces is improved or maintained comparably. From this result, it is noted that the radar range equation does not properly correct intensity in far range and adjusted radar range equation discloses limitation in the surface types that are not included in parameter estimation due to the overlap of the estimated range (section 4.5.2).

**Consistency of both sensors**

Secondly, the distribution of intensities is examined for consistency of sensors that are used simultaneously. In Table 5-2 the distribution is presented by median and standard deviation for all points in each sample set rather than sample group.

Surface	Sensor	Uncorrected data	Radar range equation	Adjusted radar range equation	Data fitting
(a) Asphalt	1	43.0/ 14.6	4.8/ 3.2	36.1/ 12.1	33.9/ 9.4
	2	34.0/ 10.9	4.8/ 4.9	32.2/ 13.9	32.7/ 11.0
(b) Road marks	1	435.0/ 151.4	54.7/ 234.6	410.7/ 120.8	372.2/ 127.1
	2	444.0/ 159.2	51.1/ 217.6	417.8/ 233.3	426.2/ 227.2
(c) Wall	1	615.5/ 341.1	2,899.0/ 571.5	1,234.7/ 456.2	319.7/ 63.0
	2	531.0/ 217.4	2,481.7/ 851.5	1,559.4/ 475.8	355.9/ 165.5
(d) Roof	1	61.0/ 129.5	1,231.3/ 611.1	209.3/ 125.9	135.8/ 67.4
	2	61.0/ 65.5	946.5/ 548.4	210.4/ 131.4	126.6/ 71.3
(e) Asphalt	1	41.0/ 9.7	4.6/ 52.0	35.4/ 8.3	32.5/ 6.4
	2	32.0/ 13.8	6.9/ 520.5	27.6/ $1.6 \times e^{16}$	34.5/ 54.6

Table 5-2: Distribution of intensities (median and standard deviation)

In the comparison of two medians by each sensor, it holds that the data fitting adjusts intensities by two sensors better than do other methods. That is, the radar range equation has serious limitation in (c) wall and (d) roof surfaces of the far range. For the adjusted radar range equation larger difference is found in (c) wall surface. The data fitting shows comparable intensity values in all sample sets. Larger difference, however, is still observed in bright surfaces: (b) road marks and (c) building wall.

**Contrast between different surface types**

This aspect is evaluated by the relative surface brightness in Table 5-2. For example, road marks and white walls have the similar brightness and are brighter than asphalt and roofs. Thus, it can be expected that (b) road marks and (c) wall surfaces have similar intensities and are higher than (a, e) asphalt and (d) roof. From this point of view the result by the data fitting has reasonable scope of intensities (asphalt - roof - wall - road marks) whereas the results by other models show unacceptable intensities in some surfaces. For example, the radar range equation shows over-estimation in further ranges, e.g. (c) building walls and (d) roofs are much higher than (a) asphalt and (b) road marks. Also, for the adjusted radar range equation it is shown that (c) building walls has higher intensity compared to other surfaces even though it is less bright than (b) road marks in reality.

**Local noise reduction of intensities**

For this evaluation, the local variation of intensities is inspected for five sample sets. Since the local variation can be derived from standard deviation in a sample patch, plots are made of ten standard deviations within each sample group in surface types (Figure 5-3). As the sample groups are also sorted by ranges of sample groups, plots can be considered as a function of standard deviation for range and this order of sample groups is same to Figure 5-2. It needs to be considered that too large standard deviation in (e) asphalt surface type may result from small sample size (less than 10) due to low point density of far ranges: the sample number of 7, 8, 9 and 10 in (e) asphalt surface.

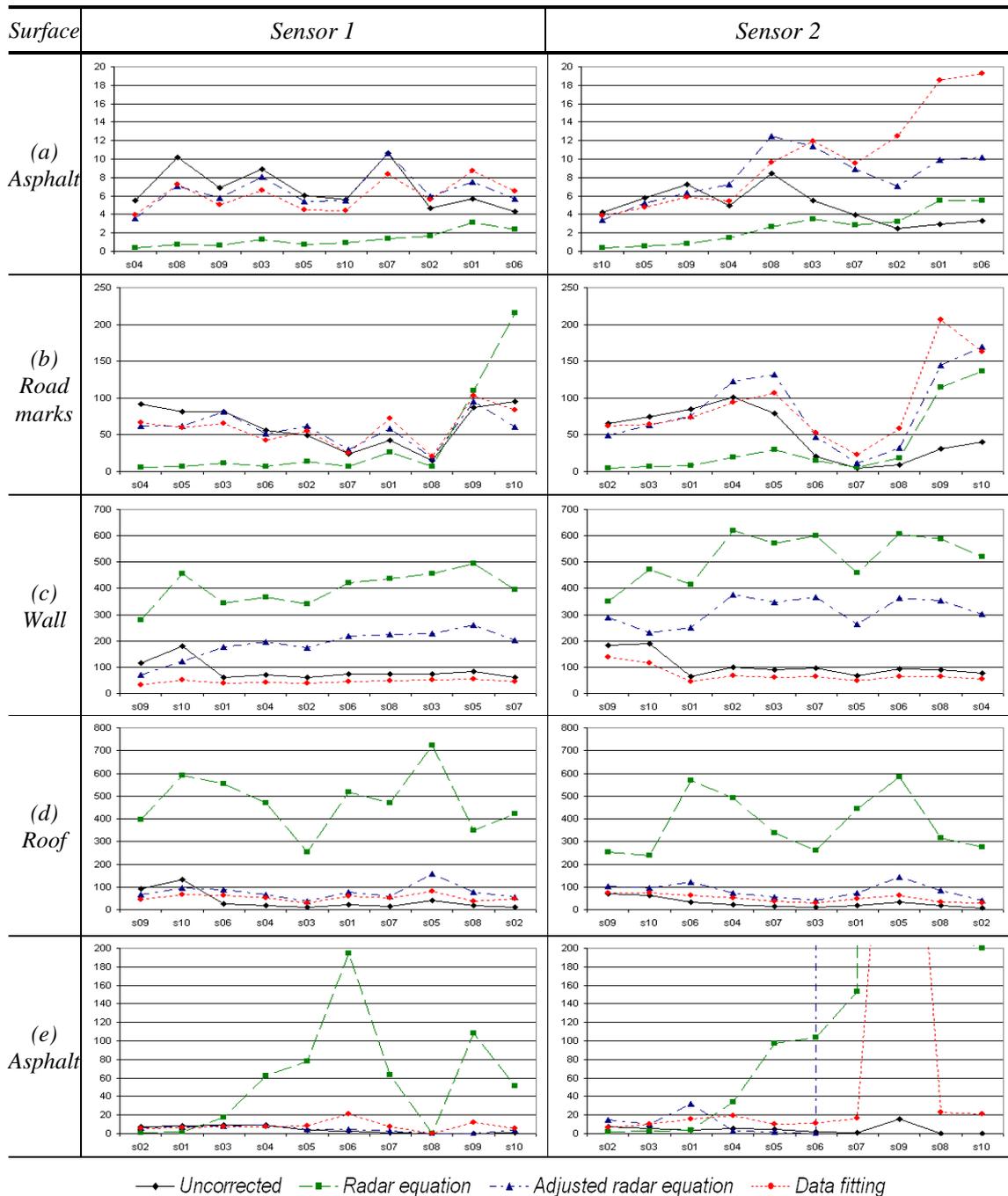


Figure 5-3: Standard deviations within sample groups

In Figure 5-3, it is shown that most lines are laid above the uncorrected data. It verifies that local noise of intensity data is not reduced in most surfaces even if outliers within sample groups have been already filtered out. When each method is investigated, the radar range equation which is presented in green longer dashed lines shows irregular variation. For example, the noise is decreased in (a) asphalt sample set of near range whereas it is irregularly increased in other sample sets of further range. That is, it is noted that this method does not have stability even though it is working better in near range. On the other hands, the results by the adjusted radar range equation (blue dashed line with a dot) and the data fitting (red small dashed line) are comparable with uncorrected in most surfaces. However, it is observed that the result by data fitting varies less than the result by adjusted radar range equation. By the same token as previous evaluation, it is noted that the result by the data fitting has smaller rates than that by adjusted radar range equation in most surfaces.

### **5.1.2.3. Conclusion**

From the above analysis on four aspects, it is revealed that the radar range equation does not reasonably work in MLS data. Indeed, it over-estimates the intensity data with increasing range so that all intensities in far range become larger than in near range. For the adjusted radar range equation it is shown that unreasonable results are produced in the sample sets which are not involved in a parameter estimation procedure. The evaluation in (a, e) asphalt and (d) building roof, which are used in parameter estimation, shows better results than in (b) and (c) surfaces in most evaluation. On the other hands, for the data fitting better results is observed compared to other methods. Therefore, it is concluded that for MLS data the empirical model based approach is more suitable than the theoretical models. The data fitting by the found trends performs best for MLS data among the methods tested.

## **5.2. Analysis by classification of surface types**

### **5.2.1. Evaluation scheme**

Up to now, the quality of methods has been statically evaluated. Thus, in this section more practical evaluation is performed with the result corrected by the best model selected in previous evaluation.

As the classification of surface types is the direct indicator on usability in laser scanner applications as well as the result provides broader information on contrast, the classification of road marks is tried from ground points which cannot be carried out by spatial geometry. For this test, another test area is used which is cropped from the same dataset. The threshold for classification is derived from test data 1 by statistical figures under the assumption that the reflectivity of surfaces is uniform in whole dataset. Finally, it is implemented by following three steps: a) filtering of non-ground points by segmentation, b) correction of intensity data, and c) classification of road marks. The result is evaluated by visual comparison for classification by uncorrected data.

### **5.2.2. Results analysis**

Based on the conclusion derived from pervious section, the result corrected by the data fitting is tested in classification of surface types (road marks). Here, the threshold is defined by the 5% and 95 % percentiles of road marks points sampled in test data 1 and is applied differently according to each sensor.

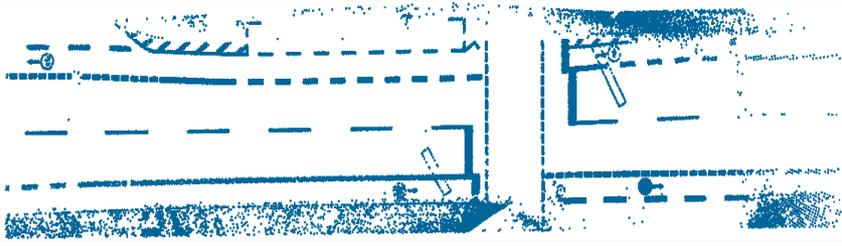
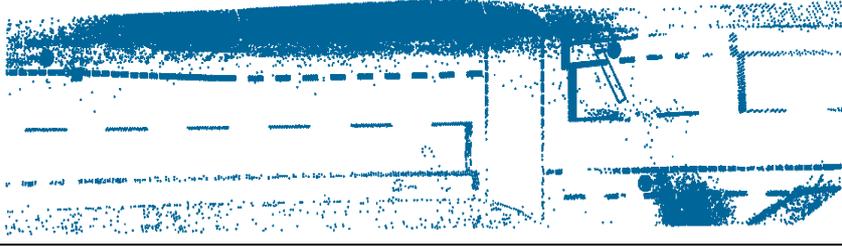
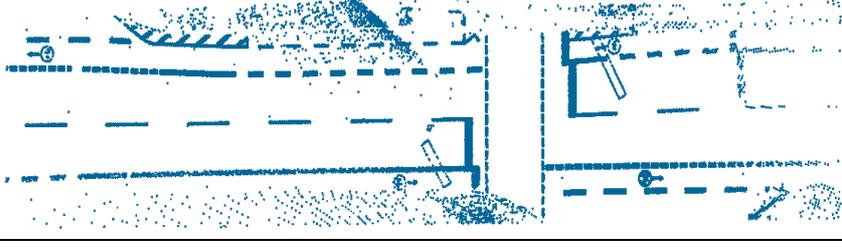
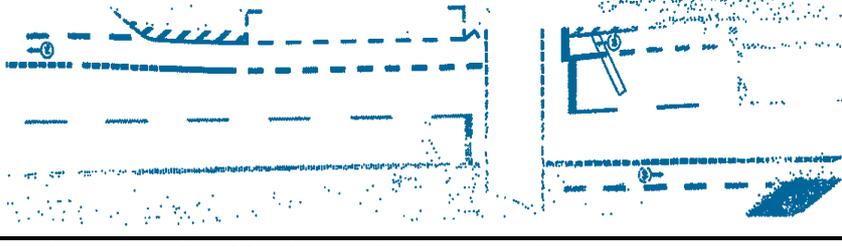
<i>Data</i>	<i>Sensor</i>	<i>Results</i>
<i>Uncorrected data</i>	1	
	2	
<i>Correction by data fitting</i>	1	
	2	

Figure 5-4: Classification result of road marks

In Figure 5-4 compared to classification by uncorrected data, it is shown that after correction the noise in most regions is reduced. At the mean time, the noise in some parts is increased such as the upper left in sensor 1. It is found that the noise is generated in the points taken from larger ranges. In addition, it was shown that the correction in near range is reliable enough to extract detailed objects (Figure 5-5). However, for more reliable results the behaviour in far range needs to be more inspected. Also, it is noted that additional polynomial models for fitting have to be tested to express the examined pattern.

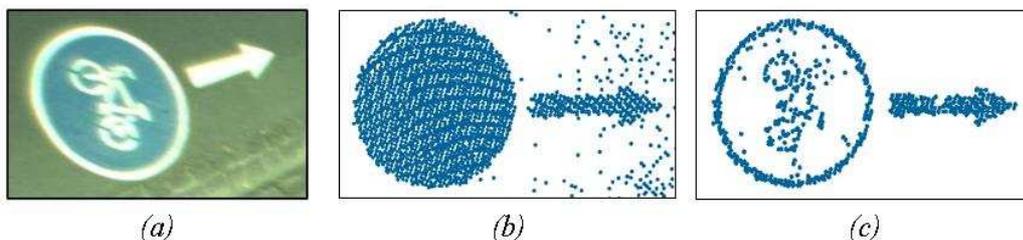


Figure 5-5: Example of a classified detailed object (a) image by digital camera, (b) uncorrected data and (c) corrected data

### **5.3. Concluding remarks**

In this chapter, detailed evaluation results of three correction methods have been presented based on statistical and practical approaches. As analyzed by the statistical approach, the radar range equation showed the decrease of intensities in near range and the increase in far range in all evaluation results. It is concluded that the radar range equation cannot be used in correction of MLS intensity data which are mostly taken within range of 50 m. It is indirectly noted that aside from ALS data sensor system properties have to be more inspected in MLS data. Moreover, the adjusted radar range equation showed limitation in surfaces that are not included in an estimation procedure. On the other hands, overall scope of range and incidence angle has to be sampled and examined in order to define parameters of two independent variables. It, however, is difficult in normal survey as long as it is not intended. Therefore, it is concluded that this method can be applied to only points with the similar geometry to sampled data rather than the overall dataset. Compared to other methods, more reliable results were presented in correction results by the data fitting. However, less accurate results were detected in surfaces with different brightness to asphalt surface which is used in modeling as well as in far range that was not thoroughly inspected due to lack of examined samples.

## 6. Conclusions and recommendations

### 6.1. Conclusions

This research aimed at developing the correction model of MLS intensity data, which have systematic biases by influencing factors such as range, incidence angle and atmospheric attenuation. In order to achieve this main objective, six research questions were formulated. (See section 1.2)

For the first question on reconstruction of scan geometry into range and incidence angle, it was presented that they can be extracted from point clouds, corresponding GPS data and offset data. It was not straightforward due to several estimations. However, range and incidence angle were reconstructed respectively from the length of vectors from sensor to target, and the angle enclosed by this vector (from sensor to target) and the surface normal, which is estimated from nearest neighbour points.

The second and third questions on the relation of intensities with respect to range and incidence angle could be answered by box plots of sample sets in section 4.5.1. In this plot, the box plots arranged by range had the drastic trends in near range. It was revealed that the two sensors have clearly different patterns. Moreover, it was enlightened surface brightness hardly affected to essential pattern of intensity and only the difference of changing degree was detected.

For the fourth question on modelling, three models were inspected. Based on the detected patterns, the simulated intensity by the radar range equation and correction result, it was noted that the radar range equation has definite limitation in MLS data that are taken from close range (mainly below 17 m). The detected relation was modelled by empirical functions: a) the adjusted radar range equation with four contribution parameters and b) the data fitting based on the relation between the range and intensity corrected by Lambertian cosine law.

The fifth and sixth questions on performance evaluation were answered in chapter 5. Three models could be assessed statistically under overall evaluation frame and practically from one of laser scanning applications. Based on analysis of results reproduced by correction models, the following conclusions are drawn:

- Aside from range and incidence angle, there are more influencing factors for intensity such as sensor system factors and signal beam divergence. They have been normally described in the full radar range equation. However, it was found that reliable results can be obtained by the function of intensity with respect to only range and incidence angle. This property will be able to be used for further improvement of correction model.
- Physical experiments on laser beam propagation have revealed that signal power is reduced by inversely quadratic function. However, in this research it was shown that this function is not proper to express intensity behaviour especially in near range. In near range, sensor system factors affect to measurement of intensity data more than do other influencing factors.

- Two sensors, which are mounted in one system platform and made by one company (Optech), showed different trends for intensities. Therefore, intensities of each sensor have to be corrected by individual correction models. In order to be able to derive one universal model for MLS with multiple laser scanners, a high accurate relative correction of the individual scanners is a pro-requisite. This kind of property has to be considered for further studies to develop correction models for MLS.

As the behaviour of MLS intensity showed different patterns in near range, correction of intensity data could not be managed by the simplified radar range equation that ignored system factors as constant factor. Moreover, correction of MLS intensity data has been difficult since the information on sensor system factors have not been provided by sensor manufacturers. Comprehensively, in this research it has been shown that the empirical model based approach can be proposed as the possible solution for lack of data on system factors in order to serve simple application such as the extraction of road marks.

## 6.2. Recommendations

This research showed that empirical functions depending on the range and incidence angle can be used in normalization of MLS intensity data. However as addressed, sampling has to be carried out by lots of manual works. Also, in normal laser scanning data it is not possible to sample a set of points including overall scope of the range and incidence angle. Moreover, less reliable results can be generated due to error propagation during estimation procedures such as surface normal, driving direction of a vehicle, and modelling of empirical functions. In order to improve the developed algorithm, it is recommended as the following:

- In this research sampling is one of the crucial procedures since samples act as representatives of whole data in data fitting. The behaviour of intensities in overall scope of range and incidence angle could not be thoroughly inspected due to lack of samples in homogeneous surfaces. Therefore, the behaviour of intensities has to be systematically examined in the overall scope of range and incidence angle which can be taken by MLS. The proposed correction model, based on Lambertian cosine law and the fitting line from a part of range, can be improved by replacing them into examined patterns.
- As mentioned, estimation of surface normal in noisy point clouds is not a straightforward process. Several algorithms have been proposed and it has been noted that a number of factors need to be considered such as point density, surface curvature and geometrical noise. Therefore, more robust estimation method of surface normal will be able to help to reduce noise due to the estimation error of surface normal.
- More sophisticated technique can be introduced to define best-fit regression line. In this research only simple polynomial functions have been used, however, it is not sufficient to define complicated patterns. Especially, linear equation currently makes sharp peaks in the intersection of two functions. Therefore, fitting with other polynomials may help to better model.

### 6.3. Further study

According to the requirement for systematic test on overall ranges and incidence angles, an experiment by reference target was organized. It aims at deriving an empirical function of intensity with respect to various ranges and incidence angles (Höfle, 2010).

In the experiment the calibrated reference target with known reflectivity behaviour was used. It was made of Spectralon® with Lambertian surface property. For scanning of target surface Optech LYNX system was used, which is the same device employed in this research. The experiment consists of two main parts for range and incidence angle. For intensity behaviour versus range, 26 measurements were performed from 2 to 52 m in 2 m step. For incidence angle 9 measurements were collected on a target rotating from 0° to 72° with 9° step at a distance of 15 m. Each measurement was collected two times, whereby the vehicle drives to and from in one direction at a distance from the target. As range and incidence angle could not be directly measured during the data collection, the places and rotated angles of the target were determined considering the relative distribution of the distance and angle.

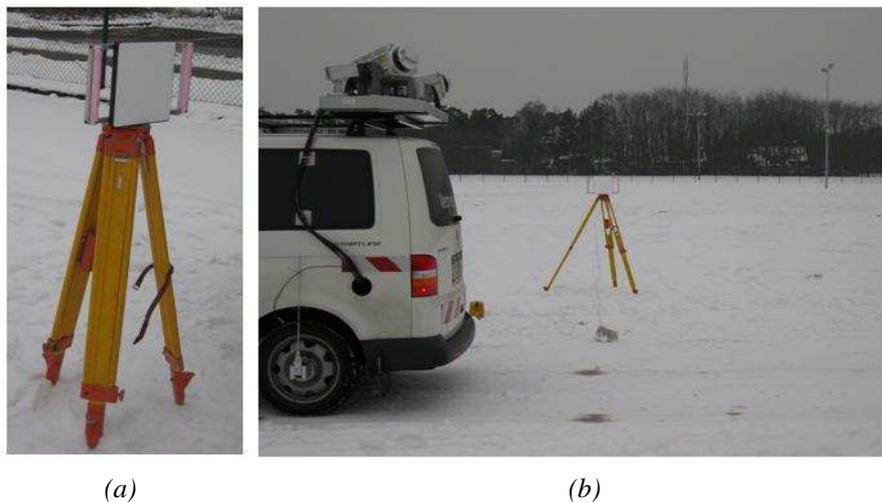


Figure 6-1: MLS intensity experiment by reference target (Höfle, 2010)

Up till now, scanning of reference target has been carried out in Rheine, Germany at 15<sup>th</sup> Jan, 2010 (Figure 6-1). From the collected data, the following scatter plots can be derived by the same way as in this research: that is, 1) Points on the target are manually sampled, 2) the geometry is reconstructed to the range and incidence angle, 3) the representative sets of points on target are made of averages of the range, incidence angle and intensity, and 4) the scatter plots are derived separately according to the experiment types (range and incidence angle). In addition, plots are divided according to driving direction (to and from) as the range and incidence angle from two ways are different.

For the distance experiment, the minimum and maximum number of points on the target was 4 in far range and 188 in the nearest range, and no point was found in 5 targets. The minimum and the maximum mean range were 1.93 m and 65.71 m, and mean intensities varied from 4.3 to 1132.7. All points contained the incidence angle of 26° ~ 43°. However, the standard deviation within a target was less than 0.7° in near range (less than 12 m) and less than 0.1° in far range (more than 12 m).

For the incidence angle experiment, two strips were inspected only from sensor 1. The minimum and maximum number of points on the target was 18 and 81, and the maximum and minimum incidence

angles were  $2.96^\circ$  and  $71.13^\circ$ . In the same token, all points included the range of 18 m ~ 21 m and the standard deviation within a target was less than 0.1 m in all cases. Mean intensities also varied from 18.7 to 256.1.

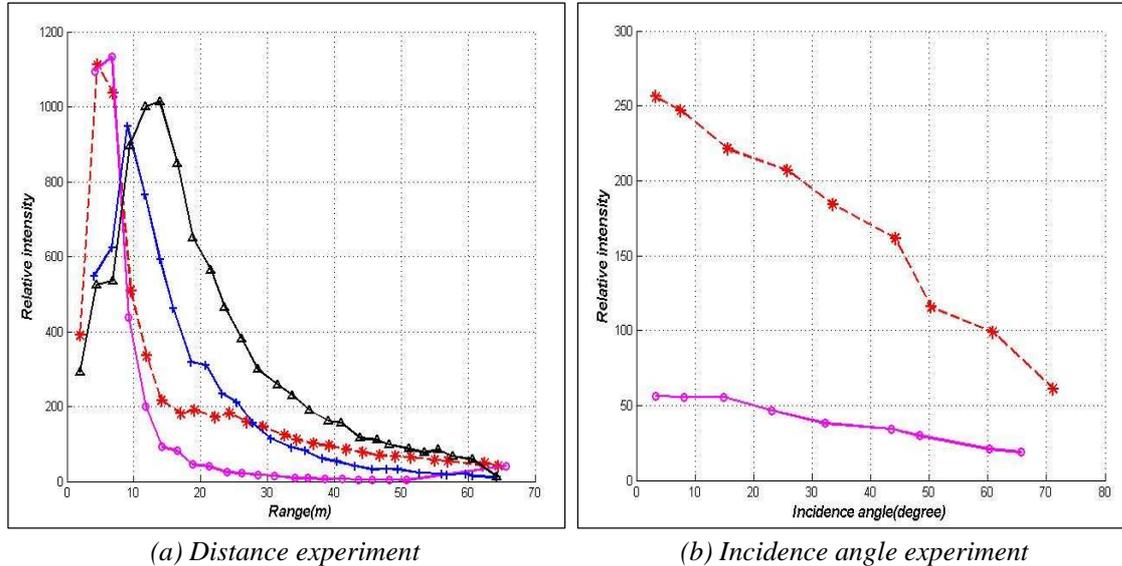


Figure 6-2: Scatter plots of experiments (red \*: strip 1 of sensor 1, magenta  $\circ$ : strip 2 of sensor 1, blue+: strip 1 of sensor 2, black  $\triangle$ : strip 2 of sensor 2)

Based on points on the target, the scatter plots are drawn where individual sensors and strips are superimposed in different colour and symbol (Figure 6-2). From the experiments it is noticed that both sensors have the similar trends for the range but the peak points are made in the different ranges. Also enlightened is that the radar range equation has definite limitation for MLS and the incidence angle must be considered in the correction of MLS intensity data. Consequently, this experiment could solve the limitation due to restricted sampling that normal scanned datasets have. Furthermore, it is anticipated that more systematic tests on fitting derived data and more detailed inspections on behaviour in near ranges (below 20 m) would be able to help the improvement of the proposed models in this research.

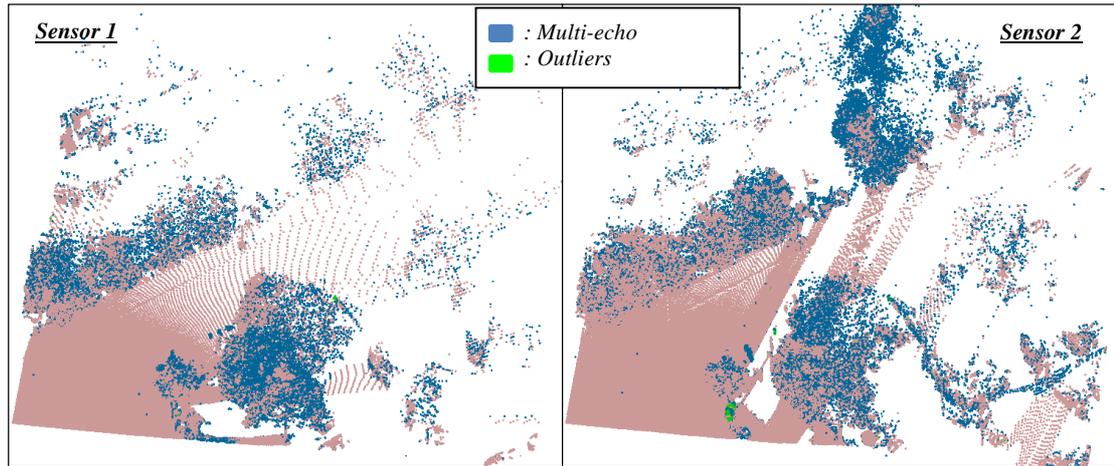
## 7. Reference

- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37: 35-46.
- Coren, F. and Sterzai, P., 2006. Radiometric correction in laser scanning. *International Journal of Remote Sensing*, 27(15): 3097-3104.
- Donoghue, D.N.M., Watt, P.J., Cox, N.J. and Wilson, J., 2007. Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment*, 110(4): 509-522.
- Eberly, D., 2008. Least Squares Fitting of Data. Geometric Tools, LLC. URL: <http://www.geometrictools.com/Documentation/LeastSquaresFitting.pdf> (Accessed December 21, 2009).
- Fernández, O., 2005. Obtaining a best fitting plane through 3D georeferenced data. *Journal of Structural Geology*, 27(5): 855-858.
- Gross, H. and Thoennessen, U., 2006. Extraction of lines from laser point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 3): 86-91.
- Höfle, B., 2010. Measurement Protocol of MLS Intensity Experiment, Internal document.
- Höfle, B. and Pfeifer, N., 2007. Correction of laser scanning intensity data: Data and model-driven approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(6): 415-433.
- Höfle, B., Vetter, M., Pfeifer, N., Mandlbürger, G. and Stötter, J., 2009. Water surface mapping from airborne laser scanning using signal intensity and elevation data. *Earth Surface Processes and Landforms*, 34(12): 1635-1649.
- Jaakkola, A., Hyypä, J., Hyypä, H. and Kukko, A., 2008. Retrieval algorithms for road surface modelling using laser-based mobile mapping. *Sensors*, 8(9): 5238-5249.
- Jelalian, A.V., 1992. *Laser Radar Systems*. Artech House, Boston London.
- Jutzi, B. and Gross, H., 2009. Normalization of lidar intensity data based on range and surface incidence angle, In: *Proceedings of ISPRS Workshop: Laserscanning 2009, Paris, France, XXXVIII(Part 3/W8)*: 213-218.
- Kaasalainen, S., Hyypä, H., Kukko, A., Litkey, P., Ahokas, E., Hyypä, J., Lehner, H., Jaakkola, A., Suomalainen, J., Akujärvi, A., Kaasalainen, M. and Pyysalo, U., 2009. Radiometric calibration of LIDAR intensity with commercially available reference targets. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2): 588-598.
- Kaasalainen, S., Kukko, A., Lindroos, T., Litkey, P., Kaartinen, H., Hyypä, J. and Ahokas, E., 2008. Brightness measurements and calibration with airborne and terrestrial laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2): 528-534.

- Kim, I.I., McArthur, B. and Korevaar, E., 2001. Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications, In: Proceedings of SPIE, The International Society for Optical Engineering: Optical Wireless Communications III, Vol. 4214: 26 - 37.
- Kukko, A., Kaasalainen, S. and Litkey, P., 2008. Effect of incidence angle on laser scanner intensity and surface data. *Applied Optics*, 47(7): 986-992.
- Lavine, M., 1991. Problems in extrapolation illustrated with space shuttle O-Ring data. *Journal of the American Statistical Association*, 86(416): 919-921.
- Luzum, B., Starek, M. and Slatton, K.C., 2004. Normalizing ALSM Intensities. Geosensing Engineering and Mapping (GEM) Center Report No. Rep\_2004-07-001. Civil and Coastal Engineering Department, University of Florida. 8 pp. URL: [http://www.aspl.ecf.ufl.edu/reports/GEM\\_Rep\\_2004\\_07\\_001.pdf](http://www.aspl.ecf.ufl.edu/reports/GEM_Rep_2004_07_001.pdf) (Accessed August 5, 2009).
- Manandhar, D. and Shibasaki, R., 2000. Geo-referencing of multi-range data for vehicle-borne laser mapping system (VLMS), In: Proceedings of 21st Asian Conference on Remote Sensing (ACRS), Taipei, Taiwan, Vol 2: 932-937.
- Mitra, N.J. and Nguyen, A., 2003. Estimating surface normals in noisy point cloud data, In: Proceeding of the nineteenth conference on Computational geometry. ACM, 322-328.
- Mount, D.M., 2006. ANN programming manual. URL: [http://www.cs.umd.edu/~mount/ANN/Files/1.1.1/ANNmanual\\_1.1.1.pdf](http://www.cs.umd.edu/~mount/ANN/Files/1.1.1/ANNmanual_1.1.1.pdf) (Accessed December 21, 2009).
- Ohio, 2008. Lidar Basics in Remote Sensing. <http://www.dot.state.oh.us/Divisions/ProdMgt/Aerial/Pages/LiDARBasicS.aspx>, The Ohio Department of Transportation (Accessed August 5, 2009).
- Optech, 2007. LYNX mobile mapper. URL: [http://www.optech.ca/press\\_LYNX\\_Release.htm](http://www.optech.ca/press_LYNX_Release.htm) (Accessed August 5, 2009).
- Pfeifer, N., Dorninger, P., Haring, A. and Fan, H., 2007. Investigating terrestrial laser scanning intensity data: quality and functional relations, In: Proceedings of the VIII conference on Optical 3D measurement tech., Zurich, Switzerland, 328 - 337.
- Pfeifer, N., Höfle, B., Briese, C., Rutzinger, M. and Haring, A., 2008. Analysis of the backscattered energy in terrestrial laser scanning data, In: Proceedings of the XXI ISPRS congress: Silk road for information from imagery, Beijing, China, XXXVII(Part B/5-2): 1045-1051.
- Phong, B.T., 1975. Illumination for computer generated pictures. *Communications of the ACM*, 18(6): 311-317.
- Shan, J. and Toth, C.K., 2008. Topographic laser ranging and scanning: principles and processing. CRC, Boca Raton.
- Vosselman, G., 2002. On the estimation of planimetric offsets in laser altimetry data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIV(Part 3A): 375-380.
- West, K.F., Webb, B.N., Lersch, J.R., Pothier, S., Triscari, J.M. and Iverson, A.E., 2004. Context-driven automated target detection in 3D data, In: Proceedings of SPIE, Automatic Target Recognition XIV, Vol. 5426: 133-143.

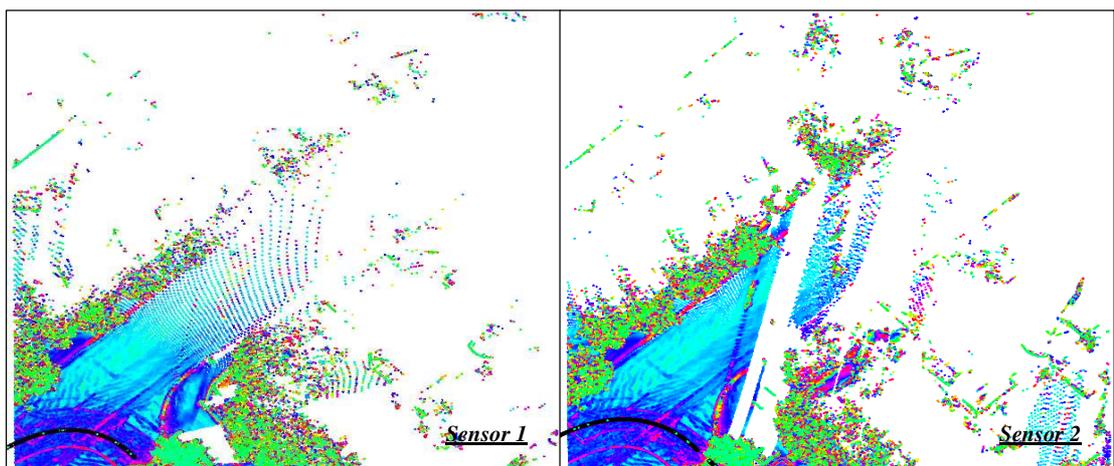
## Appendix A: Correction of test data 2 by the radar range equation

- o Pre-processing

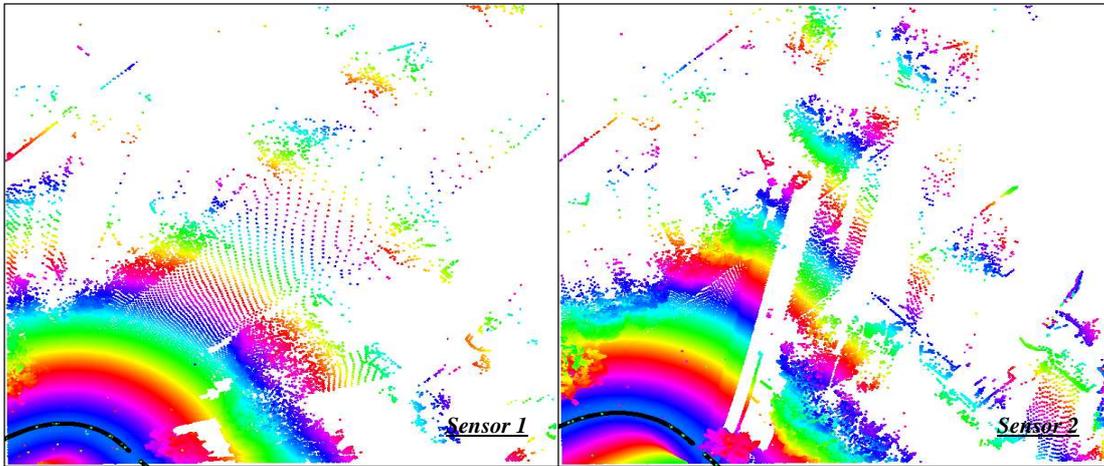


<i>Sensor</i>	<i>Multi-echoes</i>	<i>Outliers</i>	<i>Remains</i>	<i>Total</i>
<i>1 (driver side)</i>	56,341 (12.2%)	1,391 (0.3%)	405,042 (87.5%)	462,774
<i>2 (co-driver side)</i>	54,757 (11.4%)	2,327 (0.5%)	424,595 (88.1%)	481,679

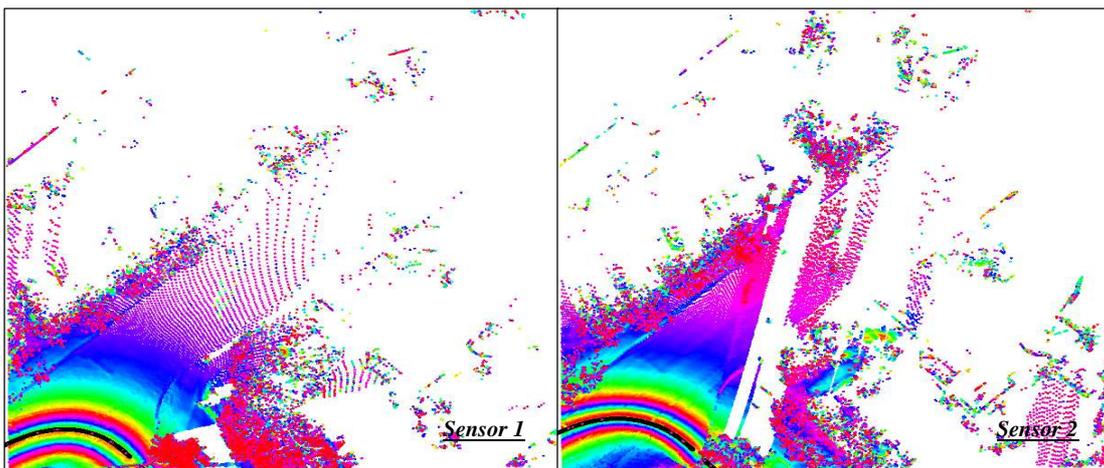
- o Surface normal



o Range



o Incidence angle



o Correction result by the radar range equation

