

博士論文

Doctoral Dissertation

Precision 3D Urban Mapping  
Using Ground and Aerial Surveillance Data

(モバイルマッピングシステムと航空測量を用いた  
都市空間高精度3次元モデリング)

Mahdi Javanmardi

ジャワンマーディ マハディ



# *Abstract*

3D urban maps serve as essential information for a wide range of applications, including autonomous vehicle positioning, drone navigation, satellite signal simulation, wireless planning, urban planning, city visualization, and disaster simulation. In this dissertation, a complete framework for the automatic urban 3D mapping is proposed which follows two main objectives: 1) Generating highly accurate and precise urban 3D point cloud; 2) 3D building map reconstruction by combining ground and aerial data, and optimizing the generated map using the satellite measurements.

A variety of applications has utilized a mobile mapping system (MMS) as the main 3D urban remote sensing platform. However, the accuracy and precision of the three-dimensional data acquired by an MMS are highly dependent on the performance of the vehicle's self-localization, which is generally performed by high-end GNSS/IMU integration. However, GNSS/IMU positioning quality degrades significantly in dense urban areas with high-rise buildings, which block and reflect the satellite signals. Traditional landmark updating methods, which improves MMS accuracy by measuring ground control points (GCPs) and manually identifying those points in the data is both labor-intensive and time-consuming. In the first part of this dissertation, a novel and comprehensive framework for automatically geo-referencing MMS data by capitalizing on road features extracted from high-resolution aerial surveillance data is proposed. The proposed framework has three key steps: (1) extracting road features from the MMS and aerial data; (2) obtaining Gaussian mixture models from the extracted aerial road features; and (3) performing registration of the MMS data to the aerial map using a dynamic sliding

window and the normal distribution transform (NDT). The accuracy of the proposed framework is verified using field data, demonstrating that it is a reliable solution for high-precision urban mapping.

Vehicle self-localization techniques, such as GNSS, visual odometry, and LiDAR scan matching can benefit greatly from 3D map of the surrounding area, especially in urban areas. However, 3D map formats such as point cloud and mesh are not suitable for these applications, because 3D map assisted vehicle self-localization requires features such as edges and surfaces of the buildings, and also the size of the 3D map should be compact. Therefore, 3D reconstruction of buildings is necessary to represent the 3D geometrical models. Accurate and automatic 3D building map reconstruction is challenging in dense urban areas, where buildings are located close to each other. If the precise 2D boundary of buildings and its roofs are available, the 3D model can be reconstructed by adding the height information to each roof segment. However, 2D maps usually provide only the rough footprint of the buildings which is not precise enough. On the other hand, with the development of laser scanning technology, airborne laser scanning (ALS) has now become readily available for large-scale city mapping, but it is still difficult to generate a 3D map using only airborne point cloud due to the difficulty in defining the original boundary of each building. In the second part of this dissertation, the precise MMS data generated in the first part is combined with the aerial point cloud to perform an accurate and precise 3D modeling. Finally, the position of the walls in the map is optimized using the satellite signal reflections.

To my dear family

تقدیم به خانواده عزیزم

# *Acknowledgment*

These sentences are my deepest gratitude to all people who supported and encouraged me during my doctoral research at the University of Tokyo, Japan.

First and foremost, I would like to express my biggest appreciation to my supervisor, Prof. Shunsuke Kamijo, for giving me the opportunity to be a member of the Kamijo Laboratory and pursuing my Ph.D. I am sincerely grateful for his moral support, steady encouragement, extraordinarily tolerant, and insightful advice. I had a great time at the Kamijo Laboratory, and I hope our collaborations will continue in the future.

I would like to give special thanks to Prof. Wei-Bin Zhang for the opportunity of the half-year research visit at the California PATH Program of the University of California, Berkeley, and all the support he has shown me throughout my research and career.

I am grateful to my Dissertation Committee Prof. Asami, Prof. Shibasaki, Prof. Sezaki, Prof. Esaki, and Prof. Naemura, for providing me extensive personal and professional guidance and teaching me a great deal about both scientific research and life.

My special thanks go to my dear brother and colleague Ehsan Javanmardi who taught me a lot of novel ideas in the research. I always admired his creativity and hard work. In addition to academic research, his advice for my various difficulties was incredibly helpful.

I would like to thank Dr. Yanlei Gu for his continuous encouragement, constructive supports during my graduate study. The discussions with him were always enjoyable. I am also thankful to our secretary Ms. Miwa for her endless support which I have greatly benefited.

I gratefully acknowledge the funding received towards my Ph.D. from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT). I also want to thank the Graduate Program for Social ICT Global Creative Leaders (GCL) of The University of Tokyo by Japan Society for the promotion of Science (JSPS) for supporting my research.

Finally but not least, I would like to express my gratitude to my parents and younger brother Sajjad for their moral support, constant understanding, and sincere encouragement.

Mahdi Javanmardi

ジャワンマーディ マハディ

# *Table of contents*

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgment</b>	<b>v</b>
<b>Table of contents</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>x</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1. Accurate and precise urban mapping framework	5
<b>Chapter 2. General Overview of 3D Urban Mapping</b>	<b>9</b>
2.1. Introduction	9
2.2. Data acquisition platforms	9
2.2.1. Mobile mapping systems (MMS)	10
2.2.2. Aerial surveillance system	11
2.3. Map format	12
2.4. Applications	12
2.4.1. Intelligent vehicles localization	12
2.4.2. Road and city asset management	15
2.4.3. Wireless telecommunication planning	16
2.4.4. Urban city analysis and planning	18
2.4.5. 3D city visualization	18
<b>Chapter 3. Registration of Mobile Mapping System and Aerial Imagery</b>	<b>20</b>
3.1. Introduction	20
3.2. Method	25
3.2.1. Extracting road markings from aerial data	28
3.2.2. Road marking extraction from filtered aerial image	32
3.2.3. Road marking extraction from MMS data	33
3.2.4. 2D registration	38
3.3. Experiment and analyses	46
3.3.1. MMS, aerial surveillance system	46
3.3.2. Experimental area	50
3.3.3. Experimental parameters	52
3.3.4. Experimental results	53
3.3.5. Discussion	55

---

3.4.	Conclusions .....	66
<b>Chapter 4.</b>	<b>Three-Dimensional Registration of Mobile Mapping System and Aerial Surveillance Data.....</b>	<b>69</b>
4.1.	Generating 3D NDT map from aerial reference.....	71
4.2.	3D MMS registration .....	72
4.3.	Experiments and evaluations.....	74
4.4.	Conclusions .....	77
<b>Chapter 5.</b>	<b>3D Urban Modeling Using Ground and Aerial Surveillance Data... 79</b>	<b>79</b>
5.1.	Introduction .....	79
5.2.	Related works.....	81
5.3.	Proposed method.....	82
5.3.1.	Building segmentation.....	83
5.3.2.	Roof segmentation.....	86
5.3.3.	Roof shape reconstruction .....	88
5.3.4.	3D building model reconstruction .....	91
5.4.	Experimental results.....	92
5.4.1.	Specification of the inputs .....	92
5.4.2.	Parameters .....	93
5.5.	Challenges .....	93
5.6.	Building wall adjustment utilizing GPS raytracing .....	95
5.7.	3D building reconstruction using MMS and ALS.....	96
<b>Chapter 6.</b>	<b>Conclusion and Future Work .....</b>	<b>100</b>
6.1.	Conclusion .....	100
6.2.	Future work .....	102
<b>Appendix.</b>	<b>GCL Social Project: Autonomous Vehicle Platform for Smart Society .....</b>	<b>104</b>
	Project description.....	104
	GCL Workshop: Safe and effective AV experiment design.....	107
	Concept, objectives and methodology of the workshop.....	107
	Program of workshop.....	109
	<i>GCL acknowledgment</i> .....	112
<b>List of Publications</b> .....		<b>113</b>
<b>References</b> .....		<b>114</b>

## *List of Tables*

<b>Table 3-1.</b> Details of the sensor platforms employed for the experiments .....	49
<b>Table 3-2.</b> Summary of the parameters applied in our experiment.....	52
<b>Table 3-3.</b> Evaluation result of survey No. 6.....	57
<b>Table 3-4.</b> The proposed framework's sensitivity to different NDT grid sizes .....	59
<b>Table 3-5.</b> Evaluation of the sensitivity of the framework to different initial window lengths .....	60
<b>Table 3-6.</b> Evaluation of the sensitivity of the framework to different numbers of required features	63
<b>Table 3-7.</b> Overall performance comparison .....	65
<b>Table 4-1.</b> Total error of the nine surveys.....	75

# List of Figures

**Figure 1-1.** Overview of the MMS registration method. Brown points show the ALS data. The middle region shows the aerial image. The black and white part shows the extracted 3D lane markings, and blue and green points show the registered MMS survey.....3

**Figure 1-2.** Overview of the 3D building model reconstruction from point cloud data. The modeling is divided into these steps: 1) Building segmentation; 2) Roof segmentation; 3) Roof reconstruction; 4) Building reconstruction. ....4

**Figure 1-3.** Flowchart of the proposed urban 3D mapping and modeling framework. ....8

**Figure 2-1.** Generalized MMS workflow [13]. ....10

**Figure 2-2.** ADAS map of the Hongo campus, University of Tokyo. This map is generated from the point cloud data acquired by MMS survey. ....13

**Figure 2-3.** Vehicle self-localization within high-definition point cloud map using 16-layer laser scanner and map matching. ....14

**Figure 2-4.** Some example of road assets to be recorded and managed using MMS survey performed regularly. ....16

**Figure 2-5.** Wireless communication ray tracing using 3D building models for base station positioning (Source: REMCOM Wireless InSite®).....17

**Figure 2-6.** The area around Tokyo tower from Google earth.....18

**Figure 3-1.** Visualized examples of the effects of the terms “accuracy” and “precision” in mapping (the red polyline is the ground-truth building footprint). ....22

**Figure 3-2.** The overall workflow of the proposed 2D MMS registration framework.....27

**Figure 3-3.** Sample road markings in the aerial images of the Hitotsubashi area in central Tokyo. Some of the problematic areas are highlighted by the red dotted rectangles. ....28

**Figure 3-4.** DSM reconstruction from ALS by 2.5D Delaunay triangulation: (a) input ALS point cloud (color denotes intensity); (b) reconstructed DSM mesh (color denotes height).....30

**Figure 3-5.** Generation of POM from the DSM: (a) Overview of the buildings’ perspective occlusion map generation; (b) original aerial image, (c) generated POM, (d) filtered result of the aerial image. ....31

**Figure 3-6.** Filtering moving vehicles from aerial image: (a) first aerial image; (b) enlarged view of some vehicles in the first image; (c) second aerial image; (d) enlarged view of the same areas in the second image; (e) vehicle filtering result; (f) enlarged view of the same areas in the filtered result, showing both filtered vehicles and a non-filtered vehicle.....32

**Figure 3-7.** Road marking extraction from an aerial image by adaptive thresholding: (a) a part of the original aerial image; (b) result of the road marking extraction process.....33

**Figure 3-8.** Road segmentation from an MMS point cloud: (a) original MMS point cloud consisting

of buildings, trees, vehicles and road signs (RGB color is derived from the camera); (b) the result of ground segmentation. The red points represent the ground, and the blue points are off-ground. ....	34
<b>Figure 3-9.</b> The effect of range and incidence angle on the shape of the laser footprint and the power of the reflected signal. As seen, as the range increase and the incidence angle decreases, area of the laser footprint increases, which degrades the quality of sampling.....	36
<b>Figure 3-10.</b> Effect of the increase of the distance to the reflected intensity value of the asphalt surface. By increasing distance, the angle of incidence inherently decreases. The red curve illustrates the curve fit to the relation between distance and intensity using a non-linear least squares fitting method. ....	37
<b>Figure 3-11.</b> Intensity calibration of MMS point cloud: (a) original intensity of the MMS point cloud; (b) calibrated intensity using the proposed method. ....	38
<b>Figure 3-12.</b> Road marking extraction from MMS point cloud: (a) original MMS point cloud (the color represents the original intensity); (b) the result of road marking extraction.....	39
<b>Figure 3-13.</b> Generation of the Gaussian distribution from the reference data: (a) estimated 1D Gaussian distribution of the sample points; (b) estimated 2D Gaussian distribution of the sample points; (c) extracted aerial road markings; (b) generated NDT map with a 2 m grid size. ....	40
<b>Figure 3-14.</b> Illustration of single laser scan and MMS survey.....	41
<b>Figure 3-15.</b> Concept of the dynamic sliding window. ....	42
<b>Figure 3-16.</b> The procedure of defining the window length for the target patch. First, the initial window around the target patch is divided into a feature grid (left). Second, the occupied cells—those containing five or more points—are calculated (middle). The window is extended by adding new patches until the number of occupied cells meets or exceeds the required feature count (right). ....	44
<b>Figure 3-17.</b> MMS system description: Mitsubishi Electric’s MMS-K320 (bottom) and the configuration of two SICK LMS-511 laser scanners and RTK GPS receivers (top). ....	47
<b>Figure 3-18.</b> Aerial surveillance system description: (a) the aircraft used for aerial data collection; (b) the sensor setup for the aerial imagery.....	48
<b>Figure 3-19.</b> The experimental area around Hitotsubashi intersection, a dense urban area of the Chiyoda-ku, Tokyo, Japan. ....	50
<b>Figure 3-20.</b> An example of number of LOS GPS satellites observed in the Hitotsubashi intersection in Tokyo, Japan. ....	51
<b>Figure 3-21.</b> Routes of the nine surveys S01–S07 on the map. The colors show the estimated GPS error reported by MMS. ....	53
<b>Figure 3-22.</b> Visual evaluation of the proposed method for survey No. 6: (a) survey route of the MMS on the aerial image (after registration); (b) enlarged view of the original MMS point cloud before registration (the red points are aerial road markings); (c) enlarged view of the MMS point cloud after applying the proposed method. ....	54
<b>Figure 3-23.</b> Estimated GPS error by MMS-K320 for each of the nine surveys. ....	55

<b>Figure 3-24.</b> Distribution of the GCPs collected in the experimental area utilizing the total station survey. The GCPs are from the corner points of road signs, which are clearly captured by both the MMS camera and the point cloud.....	56
<b>Figure 3-25.</b> The 2D error of survey No. 6 in the original data, after the landmark update, and after the proposed method.....	58
<b>Figure 3-26.</b> Generated Gaussian mixture models using different grid sizes.....	59
<b>Figure 3-27.</b> Length of the dynamic sliding window for survey No. 6.....	60
<b>Figure 3-28.</b> Visual evaluation of the proposed method: (a) survey route of the MMS on the aerial image; (b) enlarged view of the original (red) and calibrated (blue) MMS point cloud on the aerial image (the scale is 4 m); (c) enlarged view of the original (red) and calibrated (blue) MMS point cloud on the aerial image (the scale is 3 m).....	61
<b>Figure 3-29.</b> Effect of the patch length on the smoothness of the georeferenced MMS point cloud.....	62
<b>Figure 3-30.</b> Total 2D error of the nine surveys in the original data and by the proposed method....	63
<b>Figure 3-31.</b> Total 2D error of the nine surveys by the landmark update method and by the proposed method.....	64
<b>Figure 3-32.</b> Areas where the road markings were repainted after June 12, 2014 (the date of the aerial image acquisition).....	66
<b>Figure 4-1.</b> The overall workflow of the 3D MMS calibration framework.....	70
<b>Figure 4-2.</b> Generated high-resolution ground elevation profile from ALS.....	71
<b>Figure 4-3.</b> Error of the MMS laser scanning data before calibration, after the 3D calibration using the landmark updating method, and after applying the proposed methods. The marked virtual control points are relatively close to the GCPs measured for the landmark updating method.....	76
<b>Figure 4-4.</b> Visual evaluation of the proposed 3D registration method: (a) Survey route of the MMS on the aerial image; (b) Enlarged view of the original MMS point cloud on the aerial image; (c) Enlarged view of the MMS point cloud after applying the proposed method.....	76
<b>Figure 4-5.</b> Point cloud of one building after 3D registration. As can be seen misalignment removed.....	77
<b>Figure 5-1.</b> Workflow of the proposed framework.....	84
<b>Figure 5-2.</b> Building segmentation by initial 2D map. (a) Top view of the ALS point cloud. GSI 2D map is shown by red polygon. As shown, 2D map only provides a footprint of the building and it cannot represent the shape of building considering multiple roofs. (b) Extracted set of the building segment points using initial 2D map. Important boundary points of the building are missing because 2D map is not accurate enough. (c) Boundary points of the building which are mistakenly excluded from building segment. (d) Set of building points extracted using expanded 2D map with pre-defined threshold.....	85
<b>Figure 5-3.</b> Workflow of the roof segmentation.....	87

---

<b>Figure 5-4.</b> Roof segment detection using z-coordinate histogram. Red circles are possible roof segments. ....	88
<b>Figure 5-5.</b> Roof polygon reconstruction when neither first nor last point of two consequent lines are matching. Black points are points in the $\alpha$ -shape of the roof. Dashed lines are detected lines by Hough Transform. Red points are first and last points of each line. Green points are perpendicular image of the first or last point on the corresponding line. Red line is a part of reconstructed roof polygon. ....	91
<b>Figure 5-6.</b> Roof polygon reconstruction result. Black points are points in the $\alpha$ -shape of the roof. Dashed lines are detected lines by Hough Transform. Red points are first and last points of each line. Green points are perpendicular image of the first or last point on the corresponding line. Red line is a part of reconstructed roof polygon.....	92
<b>Figure 5-7.</b> 3D reconstruction steps for Shinjuku L Tower in Shinjuku, Tokyo.....	94
<b>Figure 5-8.</b> ALS point cloud for scans that its direction is perpendicular to some building edge; (a) Scanned building and flying direction; (b) Top view of ALS point cloud and reconstructed roof shape in red. ....	95
<b>Figure 5-9.</b> Building adjustment using GPS signal and ray-tracing technique .....	96
<b>Figure 5-11.</b> Workflow of the MMS 3D model reconstruction. ....	97
<b>Figure 5-10.</b> Make 3D building model utilizing ALS data and MMS data sources. In the right figure, the green part shows the MMS data and the red part shows the ALS data. ....	97
<b>Figure 5-12.</b> 3D map reconstruction by fusing MMS and ALS point clouds.....	98
<b>Figure 5-13.</b> 2.5D mesh generated from the ALS point cloud for a building in Hitotsubashi area....	98

# Chapter 1.

## Introduction

Accurate and precise three-dimensional maps are essential for many emerging applications such as autonomous vehicle positioning, drone navigation, satellite signal simulation, wireless planning, urban planning, city visualization, and disaster simulation. The accuracy and precision of the map are critical for many of these applications.

With the development of the advanced technologies such as high-resolution imagery and LiDAR sensors equipped on aircrafts and vehicles, the large-scale 3D urban mapping became available and applied for many different applications. However, automatic processing of a large amount of data provided by these high-definition sensors become a challenge for generating highly accurate and precise 3D maps.

The objective of this research is to produce a high-definition 3D map of urban area in different formats by an accurate and precise manner with a fully automated framework. The proposed framework should meet following properties:

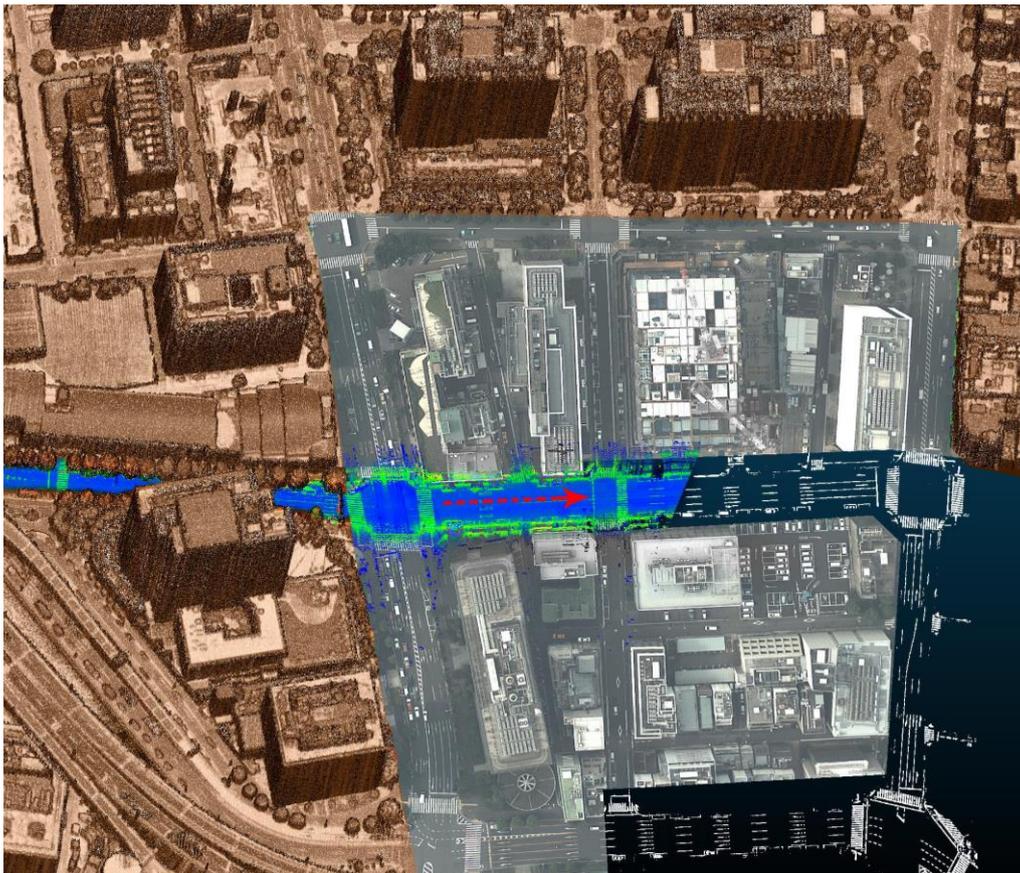
- **Automation:** The 3D map should be generated automatically with a minimum user interaction and minimized field surveys.
- **Scalability:** The framework should be able to scale for large-scale cities without losing other important properties such as automation, accuracy, and precision.
- **High-definition:** The method should be able to produce high-resolution outputs which represent the geospatial reality in the urban areas.
- **Accuracy:** The 3D information generated through the framework should have a global geospatial accuracy which is defined for different applications.

- **Precision:** The elements in the map should have high-precision which limits the uncertainty of the data, meaning that the different surveys should give the same quality data.
- **Applicability:** The method should be able to generate outputs in different formats such as point cloud, polygonal mesh and vector map to be applied for a different application.

In this dissertation, a comprehensive framework for the urban mapping will be proposed which fuses aerial surveillance data and mobile mapping. In this framework, the aerial data is taken into account for acquiring the global accuracy and the 3D geometry of the structures such as buildings, and MMS is considered for the high-definition data acquisition of the urban area from street view. The two aerial and ground-based surveillance platforms have complementary features which are effectively considered in the proposed framework. Finally, the GNSS measurements are employed to analyze and optimize the 3D position of the walls in the generated building models. Chapter 2 covers the general overview of the 3D urban mapping.

In Chapter 3, a novel framework to generate high-definition mobile mapping data is proposed which meets the above mentions requirements. The overview of the proposed framework is illustrated in **Figure 1-1**. This system contains these steps: 1) Acquiring aerial data consisting of high-resolution aerial image and aerial laser scanning (ALS) with the fusion of high-end RTK-GPS and IMU; 2) Generating high-resolution mobile mapping surveys containing camera and LiDAR point cloud using high-end RTK GPS, IMU and odometer; 3) Calibration of the aerial data with few number of ground control points (GCPs); 4) Filtering the buildings and their relief displacement in the aerial image using building occlusion map generated by perspective projection of digital surface model

(DSM) made from ALS; 5) Filtering moving vehicles from the aerial image using multiple aerial images; 6) Extracting road markings from the aerial image utilizing adaptive thresholding; 7) Generating normal distribution map from aerial road markings which is considered as the reference for the matching; 8) Subdividing each MMS survey into small patches; 9) defining a dynamic sliding window over patches to register the target patch in the middle of the window; 10) Register the dynamic window to the aerial reference using normal distribution transform (NDT); 11) Adjust the MMS trajectory and regenerate the MMS data. This method can achieve high accuracy and precision without performing labor-intensive and costly landmark update technique which requires field

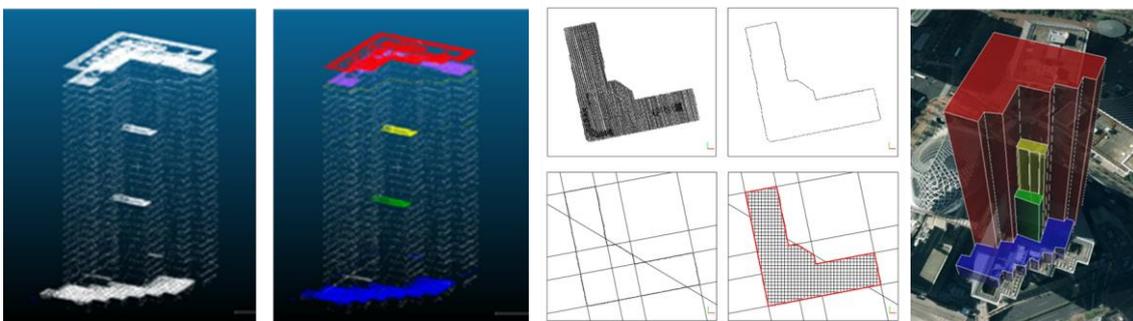


**Figure 1-1.** Overview of the MMS registration method. Brown points show the ALS data. The middle region shows the aerial image. The black and white part shows the extracted 3D lane markings, and blue and green points show the registered MMS survey.

survey for hundreds of GCP acquisition and manually picking the points in the point cloud. In most of the cases, the accuracy of the output generated using our proposed method can even outperform the traditional landmark updating.

In Chapter 4, the MMS registration framework is extended to three-dimension by generating three-dimensional road marking reference from both aerial image and laser. In the 3D framework, all registrations are extended from 2D registration to 3D by formulating the z-coordinate in the formulations. Using this framework, an accurate and precise 3D MMS point cloud is obtained.

In order to make the large size point clouds of the city with billions of points applicable for a variety of applications, an automatic 3D building modeling approach is proposed in Chapter 5 of this dissertation. The proposed method takes the combined ALS and MMS point cloud as input and generates vector based 3D building models with much less information. To generate these three-dimensional models, the buildings segmentation is first performed utilizing the 2D boundary map provided by governments. Then, a method based on the z-coordinate histogram is proposed to segment the roofs of each building. Later, the 3D shape of each roof is reconstructed by feature detection using RANSAC and the proposed roof reconstruction algorithm. Finally, reconstructed



**Figure 1-2.** Overview of the 3D building model reconstruction from point cloud data. The modeling is divided into these steps: 1) Building segmentation; 2) Roof segmentation; 3) Roof reconstruction; 4) Building reconstruction.

buildings are put in the corresponding height to generate the 3D model of each building.

**Figure 1-2** shows the 3D building model reconstruction overview. The reconstructed 3D building is optimized one more time using the satellite signal reflections to achieve the global accuracy.

## **1.1. Accurate and precise urban mapping framework**

In this chapter, a thorough framework for the urban mapping and modeling is proposed. The urban mapping suffers from some main challenges. One of the main challenges is that the aerial surveillance data do not provide street view details of the urban structures and the MMS data have a significant global error because of the GNSS/IMU integration problem in an urban area due to the blockage and reflection of the satellite signal by the tall buildings. Another main problem of the urban mapping is that the point cloud format is too big data because of its format which is not suitable for many applications. In this framework, we tackle both of these challenges to make automatic accurate and precise 3D mapping and modeling for the dense urban areas where there are a lot of tall buildings and also the distance between the buildings are less and difficult to distinguish them from each other.

In this framework, we take multiple aerial images, aerial laser scanning, MMS point cloud, MMS images, 2D boundary map and GNSS measurements as inputs and output precise and accurate 3D point cloud of the city in addition to 3D buildings models. The proposed framework fuses the aerial surveillance data and mobile mapping data to perform high-definition, high-accuracy and high-precision 3D mapping and modeling of the urban areas. In this framework, the aerial data is taken into account for acquiring the global accuracy and the 3D geometry of the structures such as buildings, and MMS which is considered for the high-definition data acquisition of the urban area from street view.

The two aerial and ground-based surveillance platforms have complementary features which are effectively considered in the proposed framework. Finally, the GNSS measurements are employed to analyze and optimize the 3D position of the walls in the generated building models. **Figure 1-3** illustrates the flowchart of the proposed framework.

As can be seen, the proposed framework is made from two main part. The objective of the first part is to generate highly accurate MMS data by registering them to aerial surveillance data. Aerial data provides 3D geometry of the buildings and roads from the top view. However, many applications require high-resolution data from the street view. Of course the MMS could a good solution for this demand but the data collected by MMS suffers from positioning error. This error makes the MMS data collected in an urban area not applicable for important applications such as intelligent vehicle's localization and navigation. To preserve the accuracy of MMS data, we proposed a framework which is illustrated in the left part of **Figure 1-3**. In this part, we take MMS data of each survey, subdivide them to small patches and register them to aerial images based on road markings extracted from both data. Since the aerial image is two-dimensional information, we cannot use it for three-dimensional registration of the MMS data. Multiple overlapping aerial images can be used to generate 3D aerial images, but the generated 3D models using this method suffers from poor triangulation. Therefore, we can use the only aerial image for the highly accurate data generation for the urban areas. Instead of using only aerial images, we take ALS data into account to generate a highly accurate 3D reference for the registration. For the registration, we propose a method using dynamic sliding window over patches of each MMS survey for geo-referencing. The output of this part is accurate MMS data and 3D road map. This information can be used in a variety of

applications including intelligent vehicles localization, navigation and perception which require highly precise maps.

In the next step, the generated point cloud information is fused with ALS to reconstruct the building models in the city. The main problem of the point cloud data is its large data size and district point structure which make it unusable for many applications. In order to make the large size point clouds of the city with billions of points applicable for a variety of applications, an automatic 3D building modeling approach is proposed which is shown on the right side of the flowchart. The proposed method takes the combined ALS and MMS point cloud as input and generates vector based 3D building models with much less information. To generate these three-dimensional models, first the buildings segmentation is performed utilizing the 2D boundary map provided by governments. Then, a method based on the z-coordinate histogram is proposed to segment the roofs of each building. Later, the 3D shape of each roof is reconstructed by feature detection using RANSAC and the proposed roof reconstruction algorithm. Finally, reconstructed buildings are put in the corresponding height to generate the 3D model of each building.

The remaining part of this document is organized as follows: Chapter 2 covers the general overview of the urban mapping framework. The 2D MMS registration framework and its details are described in Chapter 3 and different experimental results are analyzed to evaluate the proposed framework. Chapter 4 extends the 2D registration framework to 3D. Chapter 5 covers the 3D building map reconstruction from point cloud data. The conclusion and possible future extensions are discussed in Chapter 6. Finally, the Appendix covers the description of the social innovation project designed for the social ICT global creative leaders (GCL) program.

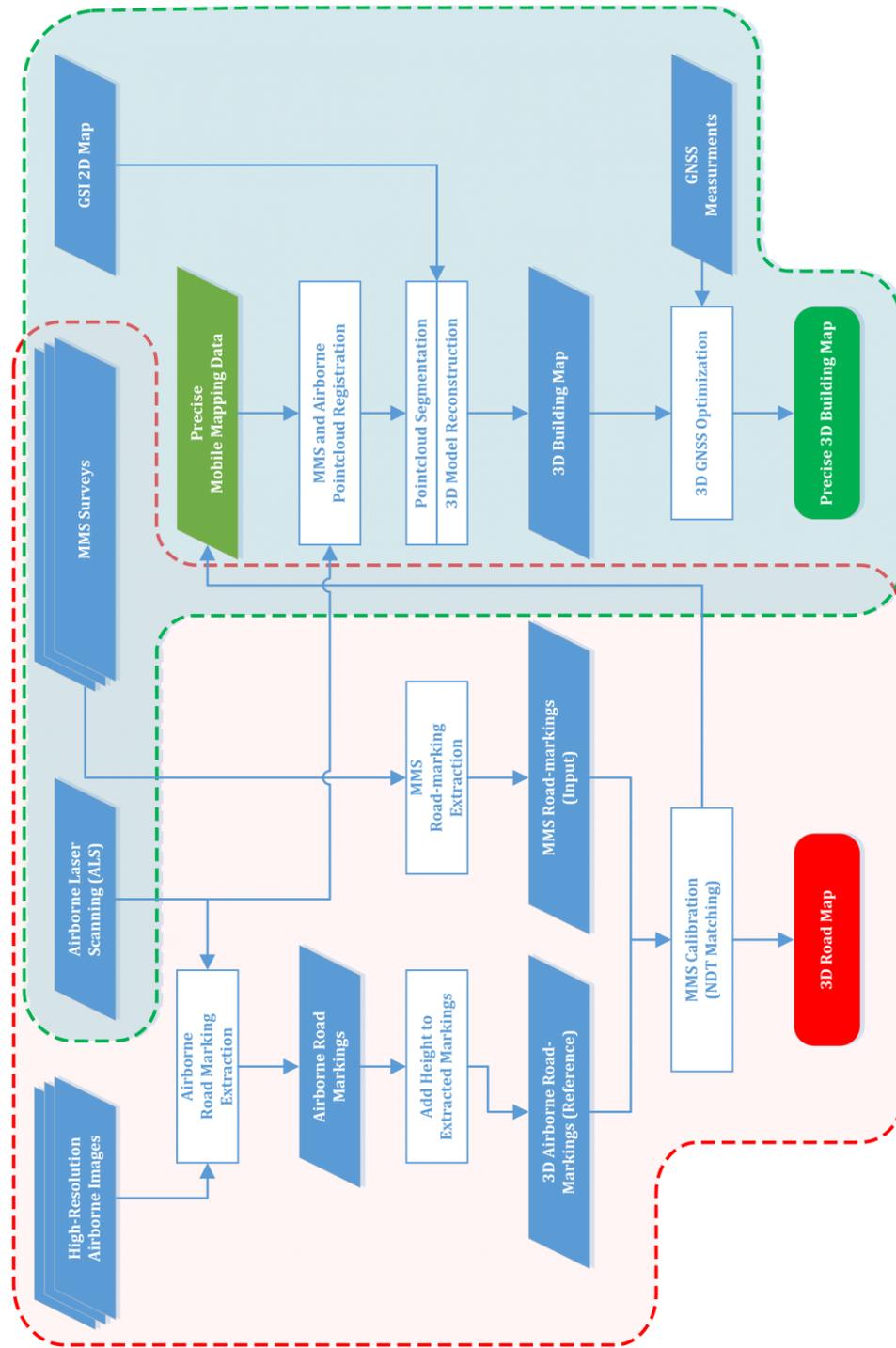


Figure 1-3. Flowchart of the proposed urban 3D mapping and modeling framework.

## **Chapter 2.**

# **General Overview of 3D Urban Mapping**

### **2.1. Introduction**

Three-dimensional map is the main provider of static environment information in the urban areas and may contain a variety of information such as roads, lane positions, road markings, intersections, road signs, buildings, bridges, vegetation and other 3D structures in the city. In the last decades, 3D city maps mainly appeared for the city visualization. However, today these 3D maps are being increasingly employed in a different formats for a wide range of applications such as urban planning [1], wireless communication planning [2], disaster simulation [3], intelligent vehicles' localization and navigation [4]–[6], road asset inventory [7], railway modeling [8], and vegetation detection [9].

Large-scale spatial information of the city is derived from different data acquisition technologies. A standard technique of the urban 3D mapping is to apply stereo vision on terrestrial, aerial and satellite imagery [10]. With the advances in the laser technologies, aerial and terrestrial laser scanners have become other popular platforms for generating 3D road and building models.

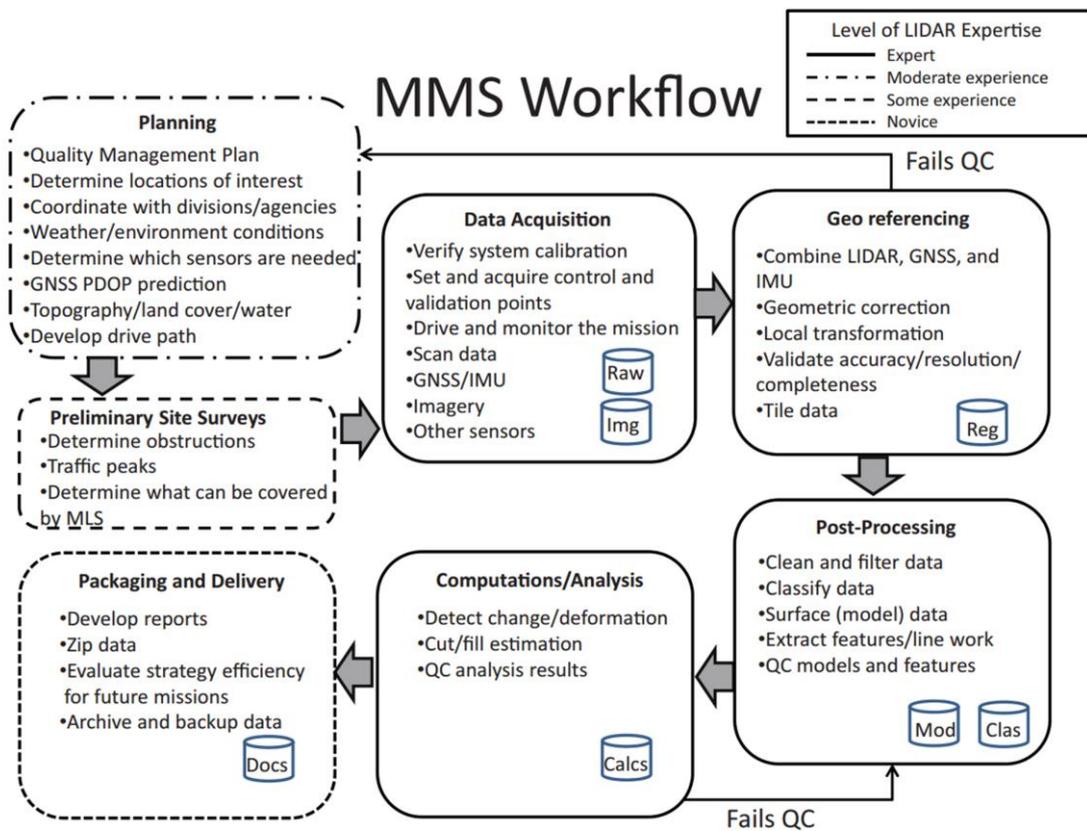
### **2.2. Data acquisition platforms**

There are a variety of data acquisition platforms for the urban mapping. In this section, we will describe the most popular mapping platforms which are MMS and aerial

surveillance system.

### 2.2.1. Mobile mapping systems (MMS)

The mobile mapping systems (MMSs) are a sensor platform mounted on the roof of a ground vehicle in combination with high-end GPS, inertial measurement unit (IMU) and odometer to rapidly capture a lot of 3D datasets essential for creating highly accurate and high-resolution representation of the roads and their surrounding environment. These systems were designed for the first time in the late 1980s when the GPS technology becomes available for public civil usages [11]. While the definition of the MMS is quite simple, the real MMS workflow contains plenty of essential factors which missing each of them leads fail in the quality control. **Figure 2-1** illustrates the generalized workflow of the MMS including middle products of each step [12]. This workflow shows the



**Figure 2-1.** Generalized MMS workflow [13].

detailed consideration of a single data collection using MMS. Traditionally, the MMS workflow consists of plenty of site surveys and experiments which require labor-intensive and time-consuming procedures.

### **2.2.2. Aerial surveillance system**

Aerial surveillance systems including aerial imagery are one of the oldest and widely used approaches for the remote sensing of the urban areas. High-resolution aerial images can provide an instant data set of a wide area of the city which is very suitable for the mapping process. Vertical aerial photos, if combined with accurate localization of the aircraft which is obtained from GPS and IMU fusion, can provide geo-referenced data which is required in a variety of applications.

Vertical aerial photos are taken with the high-resolution cameras pointed straight down. These photos contain relief displacement of the vertical structures in the urban area. If the photos are taken with considerable overlaps along the flight line and between the lines, 3D information can be generated using stereo images. However, these technique suffers from the poor triangulation. With the development of laser technologies, these aerial systems started to be equipped with LiDAR sensor for acquiring direct 3D geometry information from the high-altitude. The combination of the aerial images and the LiDAR data can provide us an accurate and high-resolution aerial data which can capture the 3D geometry of the tall buildings and their roof information which MMS cannot provide. Also, since the airplane is flying at a high altitude, its localization is much accurate compared with MMS localization. Furthermore, aerial imagery captures a large area in an instant shot which makes the relative accuracy of all pixels in an image quite accurate compared to laser point clouds.

## **2.3. Map format**

Despite the three-dimensional characteristics of the all objects in the city, the map representing them might be in 2D, 2.5D or 3D format. Moreover, 3D maps can be served in different formats such as point cloud, polygonal mesh or vector map. For deciding the dimension and format, we have to take into account all potential requirements of each specific application, and decide the benefit and advantage which each format can provide for our use.

## **2.4. Applications**

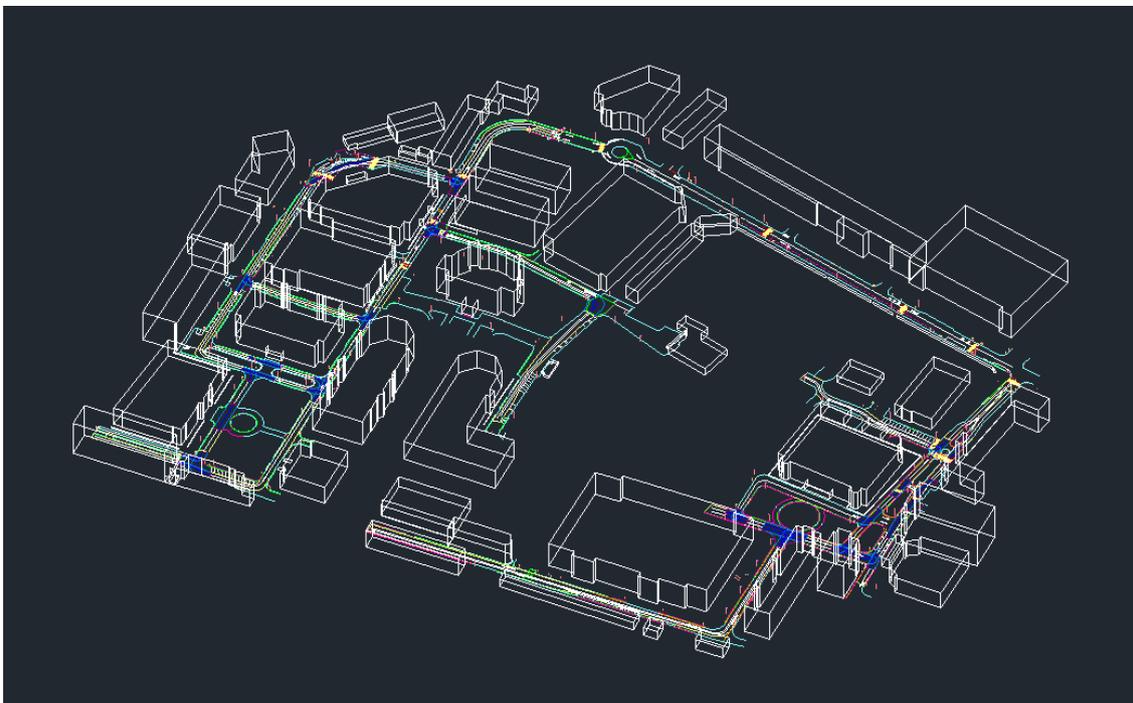
As mentioned in the introduction, three-dimensional maps have been employed for a wide range of applications. In this section, we present some of the most recent and state of the art applications of the urban 3D urban map.

### **2.4.1. Intelligent vehicles localization**

Intelligent vehicles are defined as a system that enables a vehicle to perform a collection of intelligent functions such as positioning, navigation, perception, path planning, maneuver, and collision avoidance by perceiving its environment using a variety of sensors. As the main provider of static environment information for the intelligent vehicles, the precise map is considered as a fundamental requirement for a wide range of applications.

Advanced Driving Assistance Systems (ADAS) which offer a mean to enhance active and integrated safety are one of the applications that benefit from having access to the map [13]. Detailed digital maps have been widely used for these systems and have become a “virtual sensor” integrated into the system [14]. ADAS map records a variety of information such as traffic signs, traffic lights, lane information, road slope, curvature,

and height which some of them may not be interesting for a human driver, but they offer many benefits to the ADAS applications installed on intelligent vehicles. A good example which benefits from ADAS map is the predictive adaptive front lighting system (PAFS). The light-based ADAS, which consists of swiveling and tiltable head-light to improve the illumination of the road, swivels the headlight depending on the steering angle, map information, speed and position of the vehicle, and enables a predictive and early swiveling of the headlight to an approaching curve in the map [15]. If a car knows it is on a hill and going toward downhill, it can switch to the electric power to take it to the bottom of the hill, because it has seen the destination on the map. **Figure 2-2** illustrates a sample ADAS map generated for the main campus of the University of Tokyo. The map consists of the building footprints, drivable areas, sidewalk areas, center line of the road, lane information, intersections, road markings, guardrails, poles, traffic signs, traffic lights,

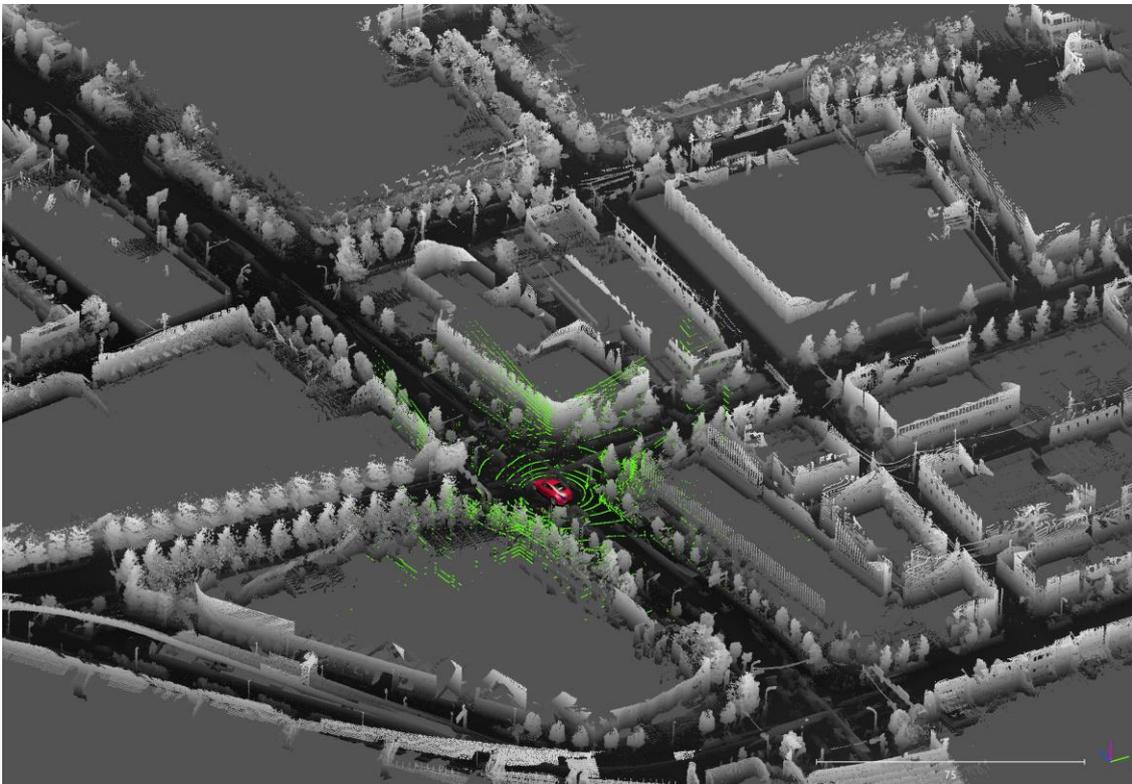


**Figure 2-2.** ADAS map of the Hongo campus, University of Tokyo. This map is generated from the point cloud data acquired by MMS survey.

and curve mirrors.

Self-localization is another crucial requirement of the intelligent vehicles which is tightly depending on the precise map. The main vehicle-based self-localization technologies such as global navigation satellite systems (GNSS), three-dimensional light detection and ranging (LiDAR), and vision-based systems can be assisted by the prior map. GNSS is proved as reliable vehicle self-positioning technology for the open sky scenario [16]. However, in the urban area with tall buildings, it suffers from the effect of multipath and non-line-of-sight (NLOS) [17]. Ray-tracing of the satellite signals reflected by buildings within a 3D building map assists GNSS to deal with the problem of signals blocked or reflected by buildings in urban areas [5], [18], [19].

One of the popular LiDAR-based localization is LiDAR odometry and mapping



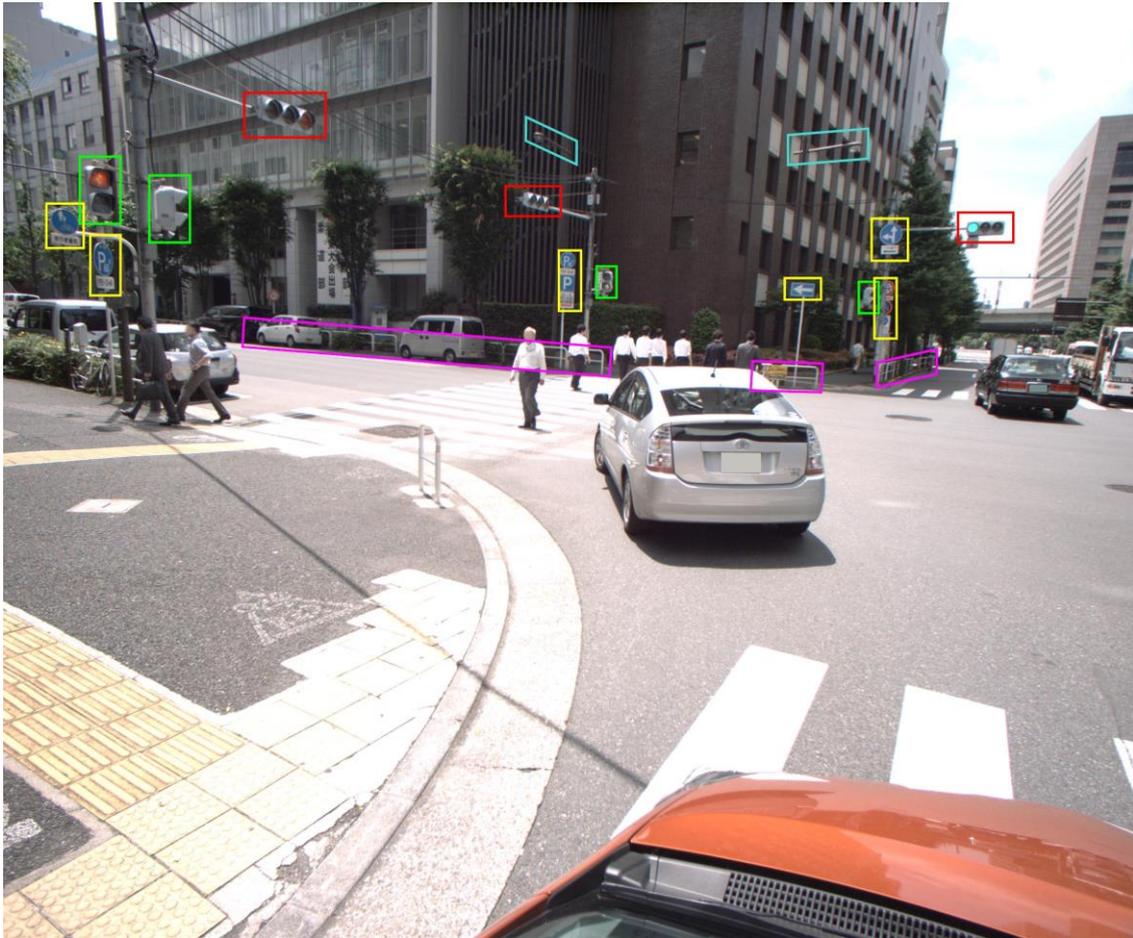
**Figure 2-3.** Vehicle self-localization within high-definition point cloud map using 16-layer laser scanner and map matching.

(LOAM). This method uses online range measurements from the LiDAR mounted on a moving vehicle to estimate the 6DOF the car. However, even the state of the art techniques using the LiDAR odometry has few meters of localization error in trajectories more than few hundred meters [20]. Another well studied LiDAR-based positioning technique for the urban area is map matching technique [21]–[23]. This method can achieve high accuracy position relative to the prior map, but the map precision and accuracy are vital for their global accuracy. **Figure 2-3** shows the vehicle self-localization within high-definition 3D point cloud map.

Vision-based localizations designed for automated urban driving have the same characteristics as LiDAR. Visual odometry suffers from the accumulative localization error [24]. Instead, many monocular and stereo vision-based techniques use a precise feature map for intelligent vehicle’s positioning [4], [25], [26].

### **2.4.2. Road and city asset management**

The government bodies and transportation agencies require managing their every road and city assets to maintain the safety standards and quality of the utilities [27]. The 3D map would also contain high-definition road and highway data which can be used for road asset management. In order to increase the road efficiency, perform the regular maintenance, improve the safety, transportation agencies are expected to record and monitor every asset in the road, and perform virtual surveys. Different efforts have been made on the national levels to give a momentum to asset management operations, and many government agencies are required or even mandated to develop a detailed road asset management plan [28]. A variety of road assets including traffic lights, road signs, and utility poles are among those items that regularly are monitored in such systems.

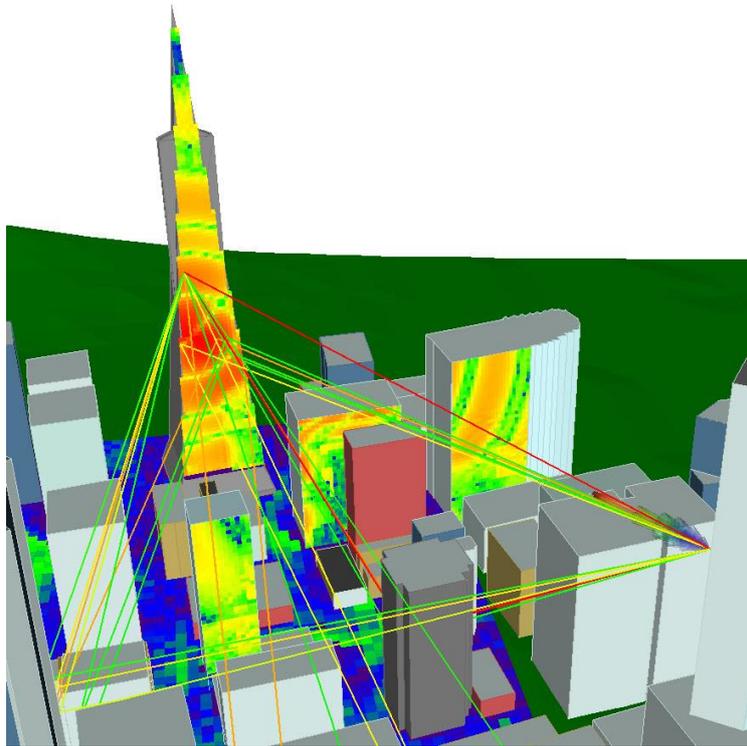


**Figure 2-4.** Some example of road assets to be recorded and managed using MMS survey performed regularly.

**Figure 2-4** illustrates an example of road assets to be managed in the map.

### **2.4.3. Wireless telecommunication planning**

Wireless communication planning of the outdoor environment is one the key requirements for the cities to keep up with new generations of wireless communication technologies and rising demands for data capacity and download speeds on cellular networks [29]. To understand the signal propagation model in the urban scenarios with dense high-rise buildings which cause shadowing by occluding the line-of-sight (LOS), channel measurement campaigns are required. To obtain enough data to accurately model the wireless channel, few months of measurement is required [30]. Instead, simulation



**Figure 2-5.** Wireless communication ray tracing using 3D building models for base station positioning (Source: REMCOM Wireless InSite®).

done by radio propagate software using ray tracing within a 3D map of the city can significantly decrease the extreme amount of time and labor required for measurements from many different spatial directions and locations. These tools enable the accurate corroboration of the field measurements by simulating the wireless propagation [29]. This application requires a large-scale 3D map of the city to perform their ray-tracings. The quality of the simulation is tightly connected to the resolution and accuracy of the generated 3D map. In addition, the simplicity of the map is required to reduce the computation time for the ray tracing. In addition, the material of the buildings facades helps to make the simulation result much more close to the reality. RabLab [31], XGtd [32] and InSite [33] are examples of signal propagation simulators. **Figure 2-5** shows the ray tracing using InSite application.

#### **2.4.4. Urban city analysis and planning**

In the field of urban planning, the term “urban models” are defined as “simplifications of reality – theoretical abstractions that represent systems in such a way that essential features crucial to the theory and its application are identified and highlighted [34].” These urban spatial models are widely being used to simulate the way cities function and convert theory into a testable form without a physical experimentation on the real world, and thus support the policy making, city analysis and planning. Over the time, theories, technologies, and methodologies have shifted from a 2D to more effective 3D models [35]. These models are used for a variety of urban planning tasks and applications such as sustainable, hazard, and policy planning, consent, shadow, lighting and visibility assessment, growth management, and site familiarization [36].

#### **2.4.5. 3D city visualization**

3D city models started evolving during the mid-1990s, due to the shift from 2D



**Figure 2-6.** The area around Tokyo tower from Google earth.

to 3D in geographic information systems, and simple 3d models were generated for the visualization purposes from the textured digital surface models (DSM) [37]. Visual 3D model of the Earth is the information extracted from the aerial and street-level data and consists of the 3D geometry of nature and human-made objects on the Earth. Google Earth and Microsoft's Bing maps are examples of the virtual earth [38]. **Figure 2-6** illustrates the area around the Tokyo Tower in Google Earth. These 3D models have a wide range of visual applications such as car navigation, virtual tourism, virtual reality, computer games and heritage preserving. Many virtual city applications do not require a high-accuracy map. However, they require an appearance looking like reality to the human eyes. Aerial stereo is advantageous for such applications, because it can provide a nice view from the top view which is suitable for virtual city and tourism.

## **Chapter 3.**

# **Registration of Mobile Mapping System and Aerial Imagery**

### **3.1. Introduction**

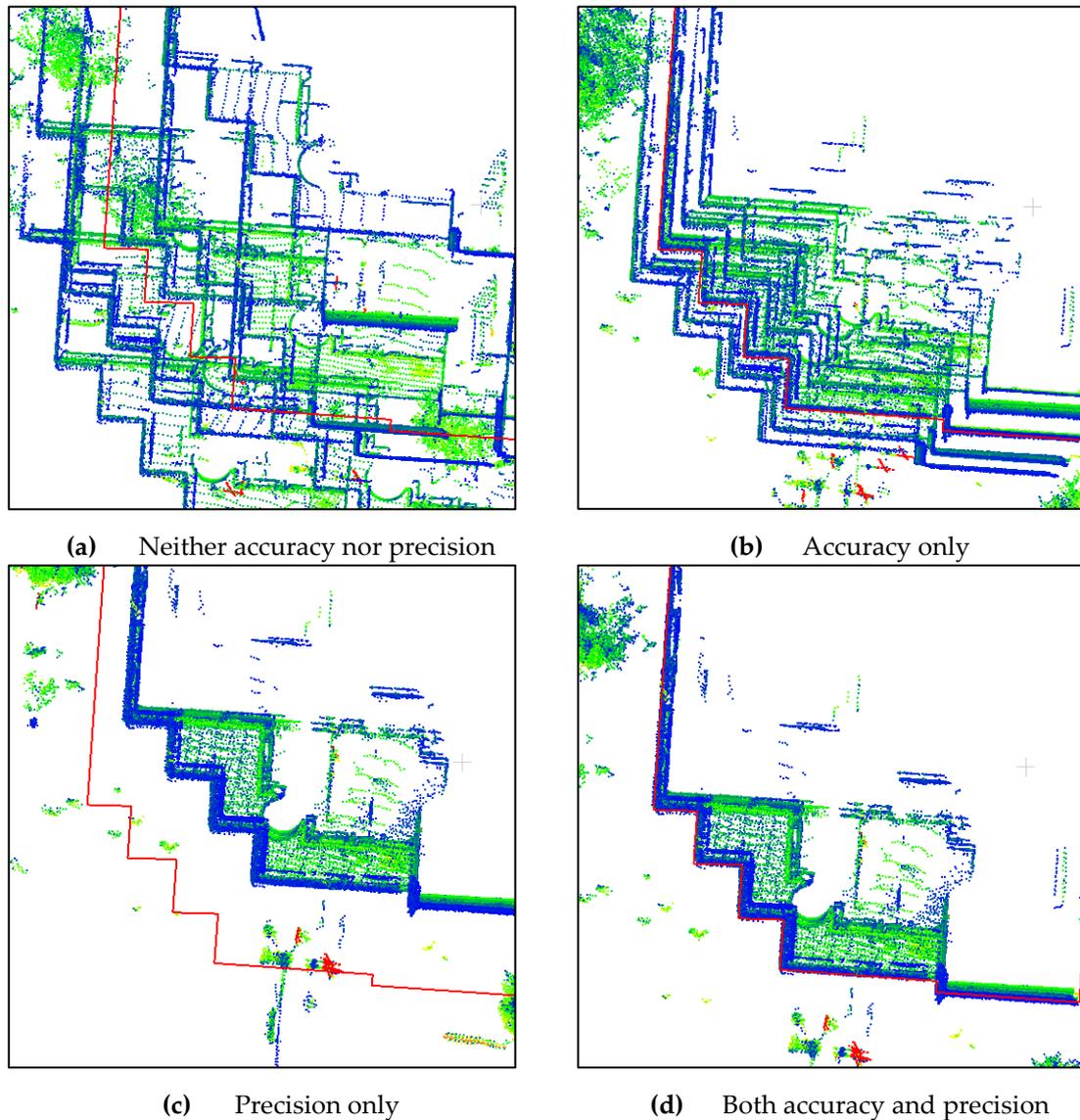
With the ability to obtain high-definition 3D environment information, mobile mapping systems (MMS) as the main urban remote sensing platform have been widely used in variety of applications such as 3D map reconstruction for intelligent vehicles' navigation and control [39]–[41], 3D city modeling [42], road asset inventory [7], railway modeling[8], vegetation detection and urban forest inventory [9], [43], [44]. As the main provider of the static environment information for the intelligent vehicles, the map made by MMS has been even considered as a “virtual sensor” [45], which enables the autonomous driving for a long mileage [46]. Advanced driver assistance systems (ADAS) which provide driver and passengers with active and integrated safety require precise sub-road details in the map, such as position and topology of pavement markings and stop lines, which is utilized for lane change assistance and lane departure warning[40], [47].

Vehicle self-localization is another application that is tightly depending on the accuracy and precision of the map. The main vehicle-based positioning technologies such as global navigation satellite systems (GNSS), 3D light detection and ranging (LiDAR), and vision-based systems can be assisted by the prior map. 3D building map helps GNSS to deal with the problem of signal blockage and reflection caused by tall buildings in the

urban area [48]. Another well studied positioning technique for the urban area is map matching technique based on LiDAR [21]–[23] and vision sensors [49]. This method can achieve high accuracy position relative to the prior map, but the map precision and accuracy are vital for their global accuracy. This method can achieve high accuracy position relative to the prior map, but the map precision and accuracy are vital for their global accuracy.

Here, an important question arises regarding how the high-accuracy and high-precision maps should be obtained in the dense urban area with tall buildings and street canyons, in a time and cost-efficient way, while no prior map exists. MMS is a valuable geo-information acquisition technology, specially designed for fast and high-definition mapping of the dense urban areas. However, its ability to acquire a high-precision 3D data is being contaminated by the vehicle’s positioning system. Self-positioning of the MMS is obtained by a tight integration of the high-end global navigation satellite system (GNSS) and inertial measurement unit (IMU) [7], [50], [51], and the 3D measurements acquired by laser scanners mounted on the vehicle are registered based on this integration result. In the urban areas where the accuracy of GNSS/IMU degrades significantly, the quality of collected data is also affected.

The map required for many of the applications listed above should be “accurate” and “precise.” **Figure 3-1** illustrates the definitions of accuracy and precision, the two important terms for the mapping. Generally, the map is reconstructed by geo-referencing multiple MMS surveys. In the urban areas, each MMS survey contains an independent error, and therefore the resulting map would be not accurate and not precise if no correction was performed (**Figure 3-1a**). In this case, the generated map cannot be used for any of applications mentioned above. Instead, if some corrections were made, but the



**Figure 3-1.** Visualized examples of the effects of the terms “accuracy” and “precision” in mapping (the red polyline is the ground-truth building footprint).

corrections were not able to limit the independent error of each survey to a certain amount, then the resulting map would be more accurate but still not precise (**Figure 3-1b**). The map after applying the landmark update with a limited number of ground control points (GCP) fit in this category. If the surveys were registered together precisely, but the global geo-referencing were not performed, the final map would be precise but not accurate (**Figure 3-1c**). Methods using simultaneous localization and mapping (SLAM) without the global reference fit in this category. In contrast, in the ideal condition, if all points of

each survey are georeferenced precisely to their correct global position, the resulting map would be accurate and precise (**Figure 3-1d**). If a map is not globally accurate, the data obtained from the map cannot be fused or integrated with a global information such as GPS data. On the other hand, if the map is not precise, acquired data such as vehicle position will not be precise as well.

Numerous studies in the past deal with the MMS registration to improve the global accuracy of the obtained data. One of the standard methods is the landmark update [8], [52], which employs the reference GCPs acquired by total station survey to update the MMS data. In this method, the reference GCPs are first measured by a field survey. Then, the reference points are handpicked in the MMS data of each survey to calculate the position correction vector (PCV). In the next step, the trajectory of the vehicle is updated by correcting the previously estimated vehicle positions using PCVs. Finally, the MMS data is regenerated with the new trajectory. Although this method can preserve some level of the accuracy, the precision of data is limited by the number GCP, where acquiring the GCP is a labor-intensive and time-consuming task. Various studies have used different methods to register multiple MMS surveys to increase the precision. Gonzalez et al. [53] proposed a method based on segmenting the highly reflective features to align multiple MMS surveys. In this approach, the global accuracy of the resulting point cloud is limited by the accuracy of the reference survey. Yao et al. [54] presented a method to register MMS point clouds to panoramic images based on a sensor constellation. Because this method relies on the relationship between the global positioning system (GPS) and a panoramic camera's position and orientation, it is not suitable for deep urban areas where buildings block the satellite signals. Aerial and ground surveillance data are two complementary data sources for high-definition 3D urban modeling [55].

Previous studies have shown that aerial surveillance data are very advantageous in maintaining the quality of ground surveys. Cheng et al. [42] proposed a method for performing automatic registration of airborne and vehicle laser point clouds based on 3D road networks and building contours. In this approach, the road networks were utilized for coarse registration, and the building contours were used for fine registration. However, because of the substantial difference between the two data, the combination of vertical and horizontal error cannot achieve better than 83 cm and 196 cm of mean and maximum error, respectively. Some other studies used airborne laser scanning (ALS) data to perform registration of terrestrial laser scanned (TLS) images [56]–[60]. However, because TLS has an entirely different error model compared to the MMS, the proposed methods cannot be applied to MMS calibration. Polewski et al. [61] proposed a method to combine the advantages of ALS and terrestrial photographs by co-registering the respective point clouds. They reported an average 2D position deviation of 66 cm in forested areas.

Another approach for using aerial surveillance data to achieve global consistency involves Lidar-based simultaneous localization and mapping (SLAM), which employs aerial images as prior information. Kuemmerle et al. [62] inserted the correspondences found between the range data and the aerial images as constraints into a graph-based formulation for SLAM. Gruen [63] utilized 169 manually extracted GCPs from the geo-referenced UAV images to update the MMS data in a 2.2 km<sup>2</sup> area. Tournaire et al. [64] proposed an image-based strategy for geo-referencing an MMS in an urban area that employed multi-view aerial images and “zebra markings” for registration. However, because their proposed method and results were preliminary, they did not perform a registration of single or multiple MMS surveys to evaluate the performance. Hussnain et al. [65] proposed a method that maintains the accuracy of an MMS by performing an

automatic registration of a rasterized point cloud to aerial image tiles. Although they reported a pixel-level accuracy, a comparison between the proposed method and manually measured road markings shows greater errors in the least accurate results. Similarly, Jende et al. [66], [67] adopted feature-matching techniques to perform registration of mobile mapping images to corresponding aerial images, but they did not report the accuracy and performance of their method. Our previous work [68] showed preliminary results of the registration of MMSs to aerial images utilizing road markings, in which the proposed framework was semi-automatic and fixed-length windows were used for the registration. However, that prior work lacked a quantitative evaluation of the georeferencing error.

In this chapter, we aim to design a fully automated framework to conduct accurate and precise mobile mapping of urban areas by employing the road markings obtained from high-resolution aerial imagery. The proposed framework achieves MMS geo-referencing by (1) extracting road features from the MMS and aerial data; (2) generating Gaussian mixture models from the aerial road features; and (3) registering the MMS point cloud to the aerial reference using a dynamic sliding window and normal distribution transform (NDT). We evaluated the accuracy of the proposed framework by applying it to the Hitotsubashi region in Chiyoda-ku, Tokyo, Japan, a dense and complex urban scene. The results demonstrated that the proposed framework provides a reliable solution for high-precision urban mapping by geo-referencing the MMS surveys with a mean error of 11.6 cm.

### **3.2. Method**

In this section, we developed a novel and thorough framework for automatic MMS geo-referencing based on high-resolution aerial surveillance data, where road features are considered as a registration reference. The proposed framework takes

multiple MMS surveys, ALS point cloud, and multiple high-resolution aerial images as input and as output provides accurate and precise MMS data.

One unique characteristic of urban areas in well-developed cities is that the road pavement typically features clear markings. Because both aerial imagery and ground-based MMS data capture these markings clearly, they are suitable features to use for aligning urban MMS surveys. This section describes how we developed an automatic framework for the geo-referencing of the MMS surveys based on high-resolution aerial surveillance data in which road features are considered as a registration reference. The proposed framework takes multiple MMS surveys, ALS point cloud, and overlapping high-resolution aerial images as input and provides accurate and precise MMS data as output. The proposed framework is performed in the following sequence. First, road markings are extracted from both aerial images and MMS point clouds. Next, a Gaussian mixture map is generated from the aerial road markings as a reference. Then, each MMS survey is subdivided into fixed-length short patches, and finally, the MMS surveys are aligned to the aerial reference based on the registration results from the road markings. To overcome the problem of roads obscured by buildings in the aerial images, which is called relief displacement and complicates road marking extraction, a perspective occlusion map is generated for each image using ALS and the image's position and orientation. In addition, overlapping aerial photographs (forward overlap) are considered to exclude moving vehicles which can be mistakenly extracted as road features. For the registration, a dynamic-length sliding window is introduced over each MMS survey to overcome distortion within the surveys. **Figure 3-2** illustrates the overall flow of the proposed framework.

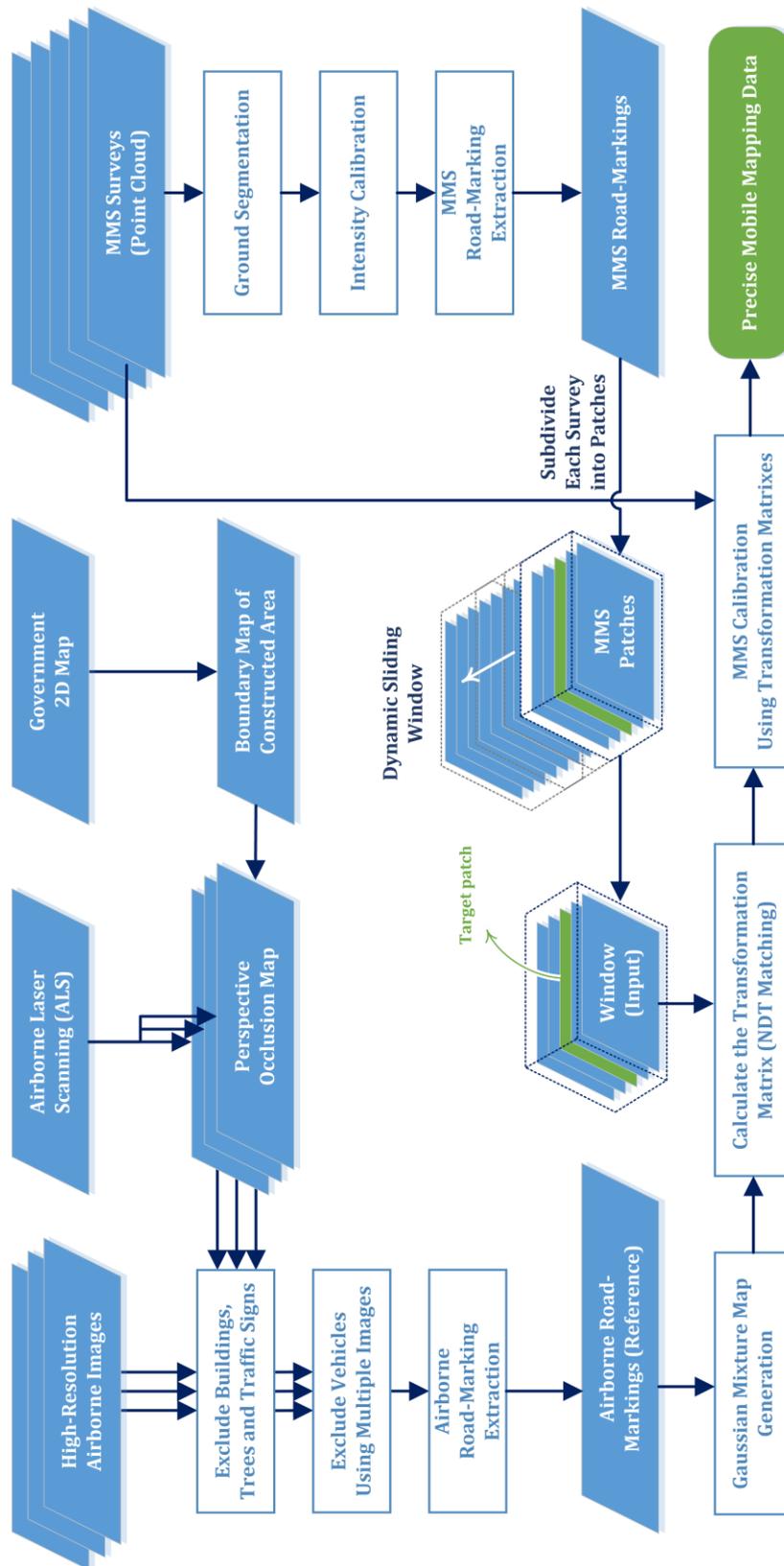
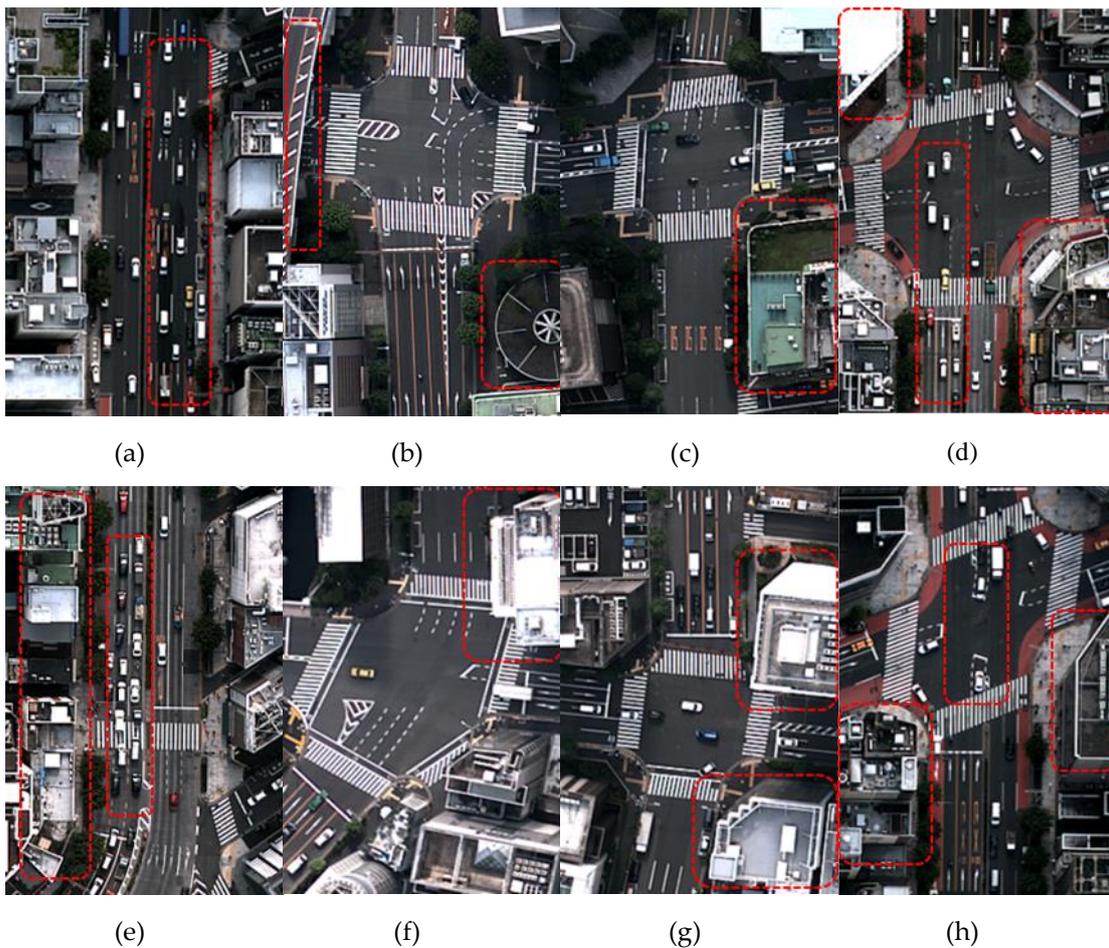


Figure 3-2. The overall workflow of the proposed 2D MMS registration framework.

### 3.2.1. Extracting road markings from aerial data

Aerial images in urban areas have some unique characteristics that make extracting road markings difficult. In urban areas, the similarity between roads and surrounding objects is significant, because roads, sidewalks, building roofs, and parking lots are made of similar materials, such as asphalt, cement, and concrete, which have a similar appearance in the images [47]. Moreover, the images often contain building edges, roof features, vehicles, and other structures that can look like road markings in the image, but are not (see **Figure 3-3a-h**). Furthermore, road obstruction by tall buildings due to the relief displacement is yet another challenge. Therefore, it is necessary to develop an



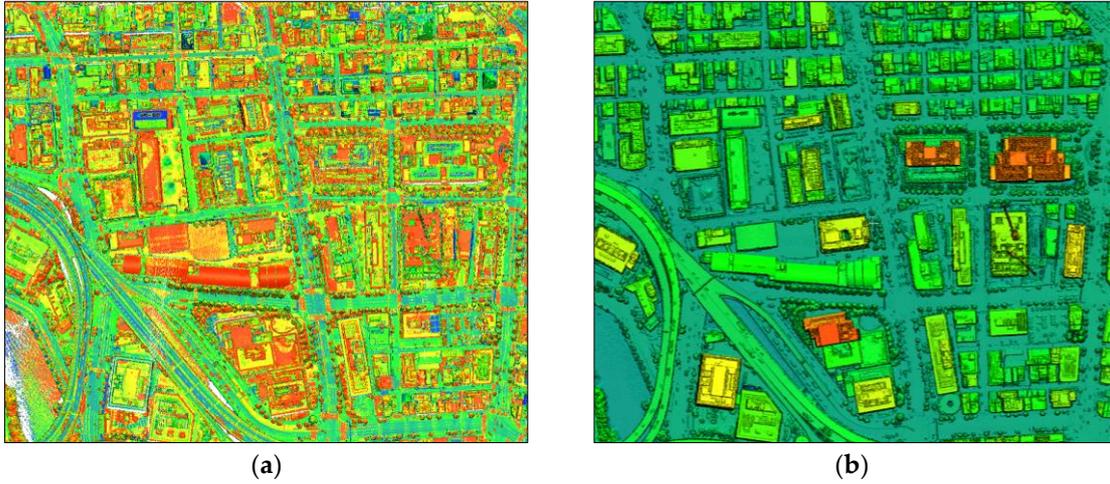
**Figure 3-3.** Sample road markings in the aerial images of the Hitotsubashi area in central Tokyo. Some of the problematic areas are highlighted by the red dotted rectangles.

automatic and thorough system that more correctly extracts road marking features for the registration to address the effects described above.

The main techniques used to extract the road markings from aerial data include road segmentation by generating a perspective occlusion map, road recovery by filtering vehicles and occluding buildings using overlapping aerial photographs, and binary image production by thresholding. These techniques are described in detail in the following sections.

### **3.2.1.1 Road segmentation with perspective occlusion maps and boundary maps**

In this section, an automatic approach that defines the road segment accurately by filtering buildings and their obstruction is proposed. In the aerial images, except for the structures at the nadir (the center of an aerial photograph), vertical objects such as tall buildings appear to lean at a certain angle. This angle increases with the distance from the principal point, which is called relief displacement. In combination with the knowledge that many building roofs have thick, sharp and white edges, two problems arise: (1) original road markings on the pavement can be occluded by the building perspectives and (2) building roof edges can be mistakenly identified as road markings. To avoid these problems, a perspective occlusion map (POM), which defines the areas occluded by buildings, is generated for each aerial image based on the digital surface model (DSM) reconstructed from ALS and the position and orientation of the aircraft at the moment of image acquisition. For the DSM reconstruction through ALS, 2.5D Delaunay triangulation is performed, which projects the ALS points in 2D on an XY plane and triangulates the corresponding points to generate a mesh structure (**Figure 3-4**). Then, the points higher than a certain threshold are considered when generating the POM. This threshold should be an above ground height to ignore ground points and vehicles on the



**Figure 3-4.** DSM reconstruction from ALS by 2.5D Delaunay triangulation: (a) input ALS point cloud (color denotes intensity); (b) reconstructed DSM mesh (color denotes height).

road surface when generating the POM.

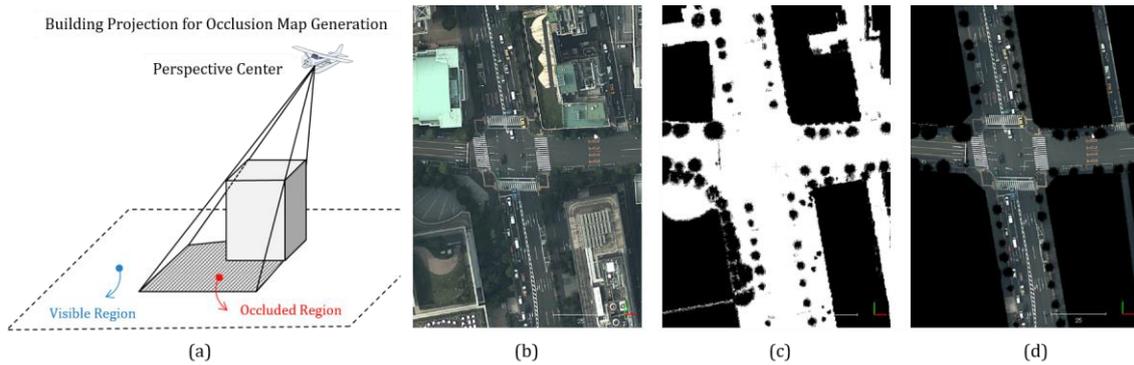
Consider a set of triangles connected by their common edges or corners that representing the DSM mesh. To reconstruct the POM, the DSM should be rasterized, which is done by projecting the 3D vertices of each off-ground triangle onto the 2D raster using perspective projection and filling all the pixels covered by the projected 2D triangles (**Figure 3-5a**). The resolution of the POM is defined as being equal to that of the aerial image. The aircraft's position  $(c_{x,y,z})$ , orientation  $(\omega, \phi, \kappa)$ , and the camera's parameters define the projection transformation. To define the 2D projection  $(x, y)$  of each vertex  $(X, Y, Z)$ , we first define the position of the vertex with respect to the camera's coordinate system  $(d_{x,y,z})$  using the following transform:

$$\begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{bmatrix} \begin{bmatrix} \cos \phi & 0 & -\sin \phi \\ 0 & 1 & 0 \\ \sin \phi & 0 & \cos \phi \end{bmatrix} \begin{bmatrix} \cos \kappa & \sin \kappa & 0 \\ -\sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \left( \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} \right), \quad (1)$$

$$x = x_0 - \frac{f}{d_z} d_x, \quad (2)$$

$$y = y_0 - \frac{f}{d_z} d_y, \quad (3)$$

where  $f$  is focal length;  $(x_0, y_0)$  indicates the image coordinates of the principal point;



**Figure 3-5.** Generation of POM from the DSM: (a) Overview of the buildings' perspective occlusion map generation; (b) original aerial image, (c) generated POM, (d) filtered result of the aerial image.

and  $c_{x,y,z}$  presents the 3D position of the camera installed on the aircraft. **Figure 3-5b** illustrates the generated POM. It can be seen in **Figure 3-5c** that the POM can filter all buildings and their relief displacements effectively. In order to have clearer road segments, we have also used the two-dimensional boundary map of the buildings provided publicly by The Geospatial Information Authority of Japan (GSI). The final result of the road segmentation is shown in **Figure 3-5d** where the generated map filters all the buildings and their relief displacements effectively.

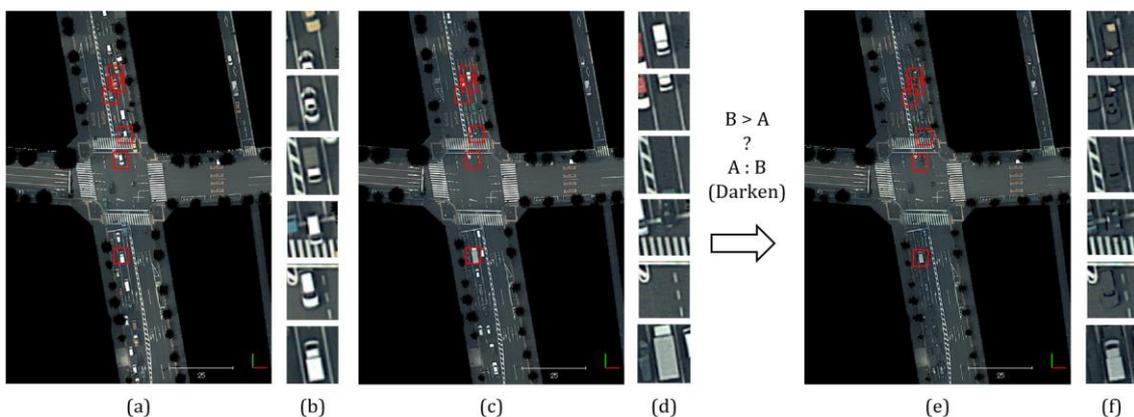
### 3.2.1.2 Filtering moving vehicles from the road

In the road marking extraction from aerial images, vehicles painted in light colors can be mistakenly considered as road markings. Since the objective of the road marking extraction is the registration between ground and aerial data, we do not require a perfect road marking extraction. The extracted road markings are considered as features for the registration. Therefore, in areas with few road marking features, missing a lane marking or misdetection of a vehicle as a road marking may have a significant effect on the registration. On the other hand, in areas with plenty of features such as intersections with zebra crossings, missing few features does not affect the matching quality. To obtain the

best vehicle filtering result, learning methods are suggested. However, we did not have access to a large and proper data set for the training. Therefore, we applied a simple method which uses multiple aerial images and filters the moving vehicles. If a vehicle is detected in the first image and it is moving, it will be more likely not to see the same vehicle in the second image and so on. The proposed solution takes multiple images and compare the total of their RGB channels, then choose the lowest value. Basically, vehicles in the aerial images are between the lane markings or may be on the zebra crossings. This simple method not only can filter all the vehicles between the lane markings but also filters all vehicles with light colors. However, the parked vehicles cannot be filtered, and vehicles with a dark color can filter a part of zebra crossings. Although this simple solution is not perfect, it is sufficient enough for our application. **Figure 3-6** illustrates these steps.

### 3.2.2. Road marking extraction from filtered aerial image

The road markings extracted from the aerial image are considered as references for the geo-referencing of the MMS data. To obtain more features for the registration, all

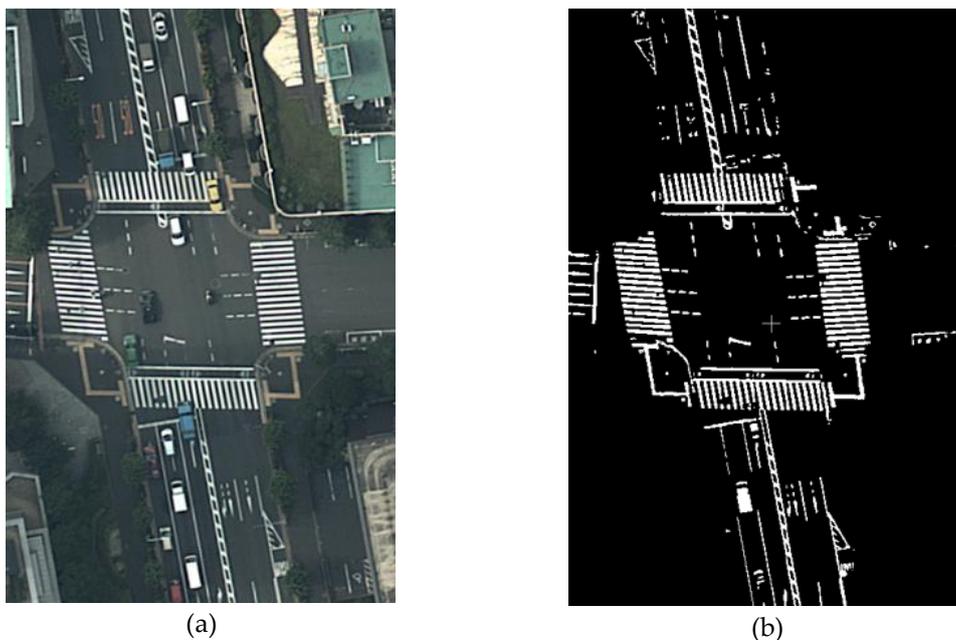


**Figure 3-6.** Filtering moving vehicles from aerial image: (a) first aerial image; (b) enlarged view of some vehicles in the first image; (c) second aerial image; (d) enlarged view of the same areas in the second image; (e) vehicle filtering result; (f) enlarged view of the same areas in the filtered result, showing both filtered vehicles and a non-filtered vehicle.

kinds of road markings and even signs in the sidewalks are considered based on the intensity contrast between the markings and the road surface background. After a level adjustment of the image, the adaptive Gaussian thresholding is performed for the road marking extraction. **Figure 3-7** illustrates an example of extracted road markings from the aerial image in the bitmap format, where white pixels are the extracted road markings and black pixels are the background.

### 3.2.3. Road marking extraction from MMS data

Next, we need to extract similar road markings from the MMS point cloud data. Similar to the procedure adopted for the aerial images, the first step is road segmentation. Then, the extraction process is performed on each road segment. In addition to precise distance, most Lidar scanners installed on the MMS also record the intensity, called the reflective luminance, which is a measure of the return strength of the laser pulse that generated the point. The intensity value enables the differentiation of different materials such as asphalt and pavement markings. The main flow of the process to extract road

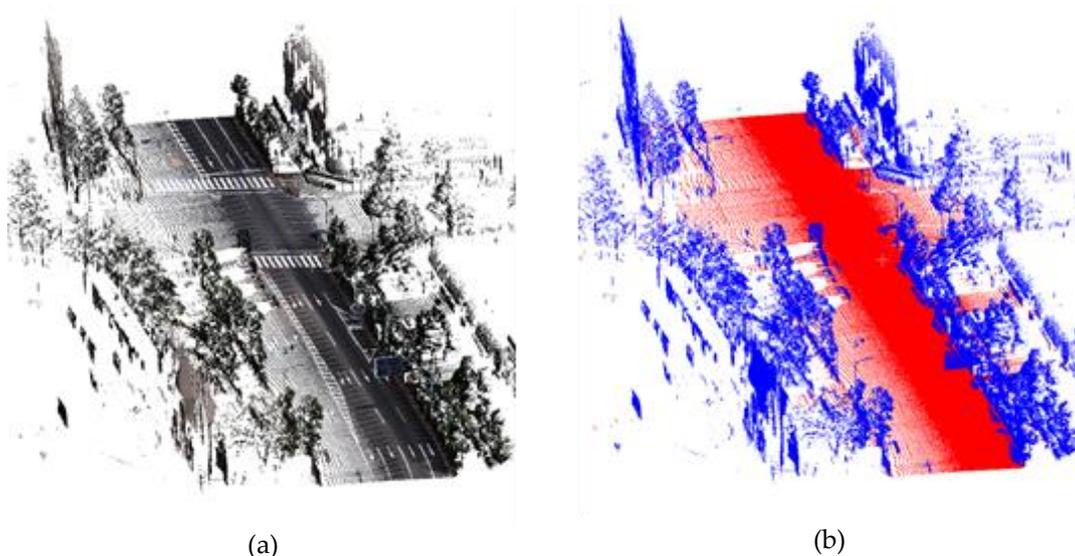


**Figure 3-7.** Road marking extraction from an aerial image by adaptive thresholding: (a) a part of the original aerial image; (b) result of the road marking extraction process.

markings from the MMS data includes road segmentation from the MMS point cloud using a cloth filter simulation, intensity calibration based on distance and incident angle, and road marking extraction by intensity thresholding.

### 3.2.3.1 Ground segmentation of the MMS point cloud

Ground segmentation from ground and airborne laser point cloud is a well-researched topic [8], [69]–[71]. Curb-based road surface segmentation is a popular method for ground point extraction from MMS data [72]. However, both road surfaces and sidewalks include signs and features suitable for matching. Because pavement markings on both roads and sidewalks are extracted from aerial images, they should also be considered in the MMS data. Therefore, the ground segmentation from MMS should be able to extract both roads and sidewalks while excluding buildings and other ground structures. For this purpose, a method based on the cloth filter simulation [73] is used for ground segmentation. **Figure 3-8** shows the result of ground segmentation in a part of MMS point cloud.

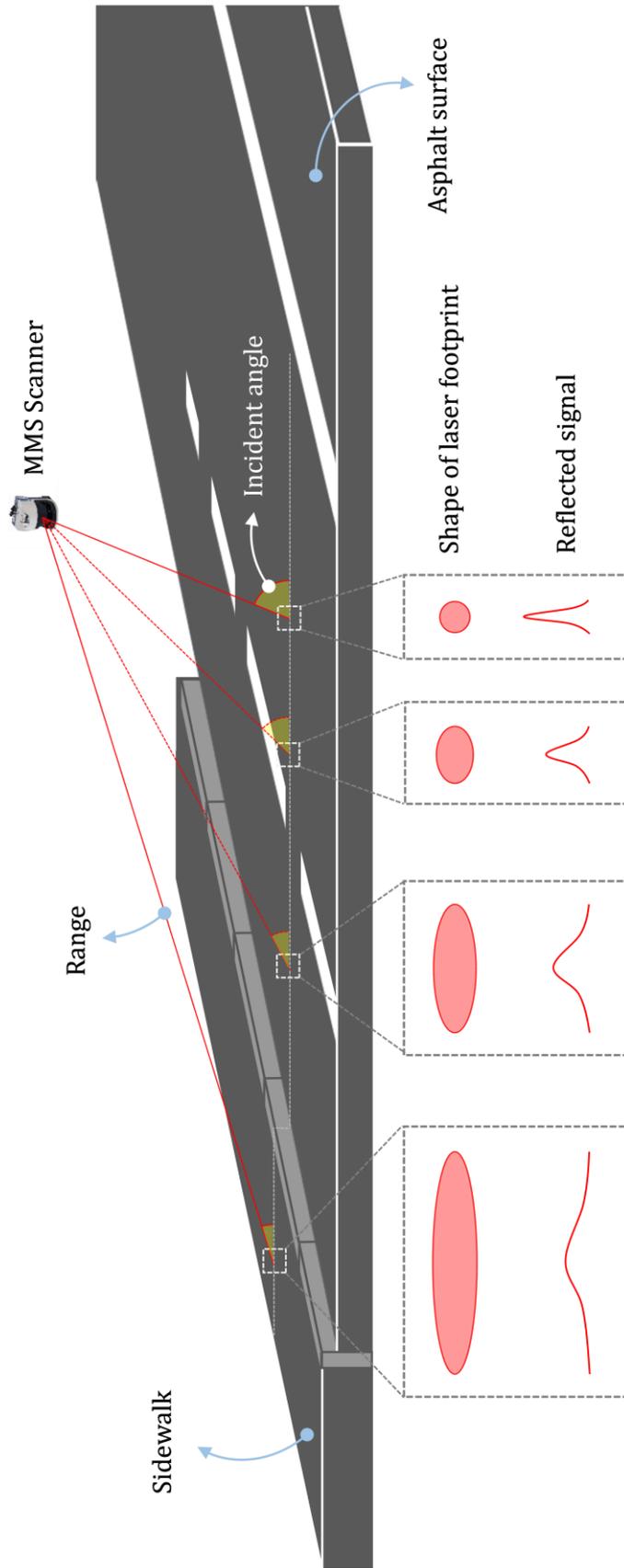


**Figure 3-8.** Road segmentation from an MMS point cloud: (a) original MMS point cloud consisting of buildings, trees, vehicles and road signs (RGB color is derived from the camera); (b) the result of ground segmentation. The red points represent the ground, and the blue points are off-ground.

### 3.2.3.2 Intensity calibration based on the distance and incidence angle

The laser scanners installed on MMS can capture the received signal strength indication (RSSI) of the returned signal (i.e., intensity/remission), which can be used to identify the material of the reflecting surface. Various research groups have used this value to differentiate between asphalt and painted parts of the road surface [72]. However, the reflected laser signal is significantly affected by the scanning geometry, mainly the distance and the laser incidence angle to the target surface [74]–[77]; therefore, it cannot be directly used for road marking extraction. In MMS scanning, where the distance between the scanner and target is relatively close, and the target surface is larger than the footprint of the laser beam, the range dependence can be expressed as  $1/R^2$ , where  $R$  is the range [74]. Another important parameter is the incidence angle. The laser footprint increases significantly as the incidence angle increases and, therefore, the quality of data is affected [78]. Although the effects of distance and incidence angle on laser intensity would ideally be modeled individually, different laser scanners from different manufacturers can show totally different responses [76]. The transmitted energy, intensity bit depth, amplification of low-reflectivity surfaces, and aperture size are some instrumental factors that affect the intensity measurements and differ between manufacturers.

In this research, the intensity value is used to differentiate the painted area of the roads and sidewalks. One of the characteristics of the road surface is its flatness. As shown in **Figure 3-9**, if the flat surfaces such as road and sidewalk are considered, the incidence angle of the laser beam is increased as the range increases. In this case, we can consider the intensity change of material only based on the distance of the measured point from the MMS scanner, which inherently contains the effect of the incidence angle.

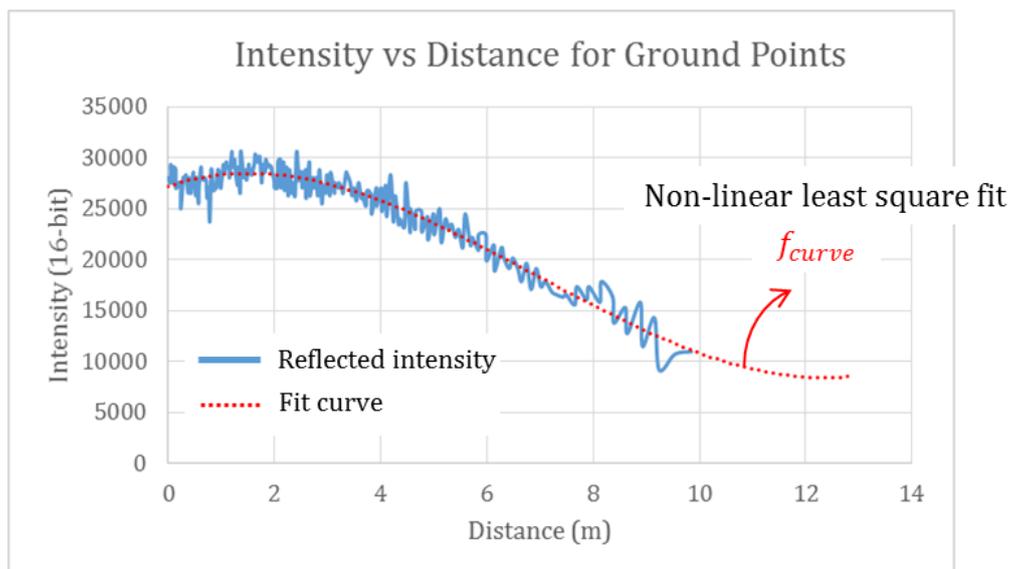


**Figure 3-9.** The effect of range and incidence angle on the shape of the laser footprint and the power of the reflected signal. As seen, as the range increase and the incidence angle decreases, area of the laser footprint increases, which degrades the quality of sampling.

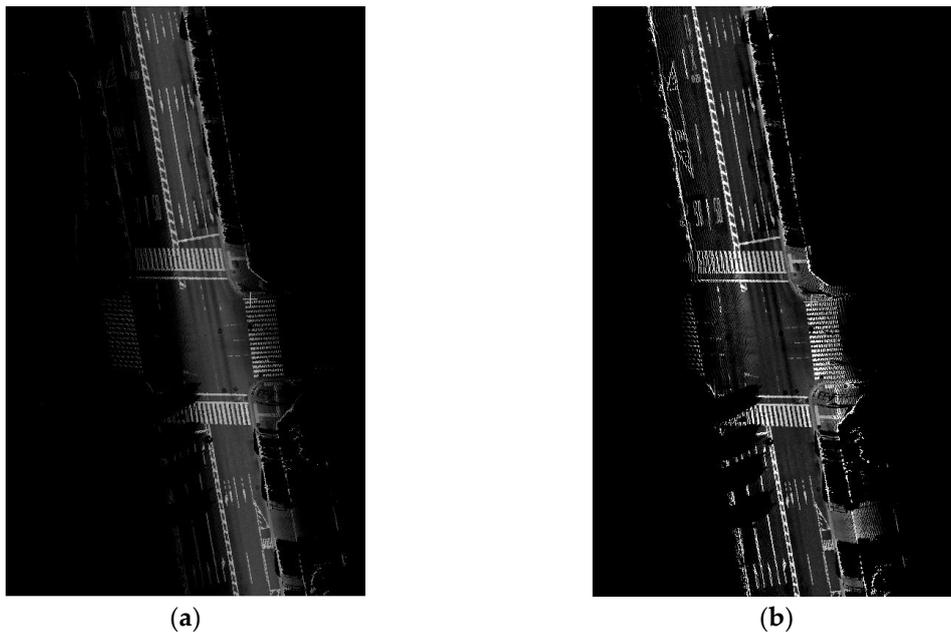
**Figure 3-10** shows the effect of distance (consisting the effect of incidence angle inherently) on the reflected intensity of the asphalt surface of an experimental road using laser installed on the MMS. In the proposed method for the calibration of these two effects, the non-linear least square curve fit is employed to model the effect of a change in range and incidence angle, and then the fit curve is used to calibrate the intensity using following empirical model:

$$I_c = \begin{cases} R < R_1 & (I_M \times I_{ref})/f_{curve}(R) \\ R_1 < R < R_2 & (I_M \times I_{ref})/I_m, \\ R > R_2 & ignore \end{cases} \quad (4)$$

where  $I_{ref}$  and  $I_c$  present the reference and calibrate intensity;  $I_M$  and  $I_m$  indicate the maximum and minimum reflection of the asphalt surface;  $R$ ,  $R_1$  and  $R_2$  are the reference range, first and second predefined range threshold consecutively; and  $f_{curve}$  indicates the fit curve acquired from least square fit. The result of intensity calibration is illustrated in **Figure 3-11**.



**Figure 3-10.** Effect of the increase of the distance to the reflected intensity value of the asphalt surface. By increasing distance, the angle of incidence inherently decreases. The red curve illustrates the curve fit to the relation between distance and intensity using a non-linear least squares fitting method.



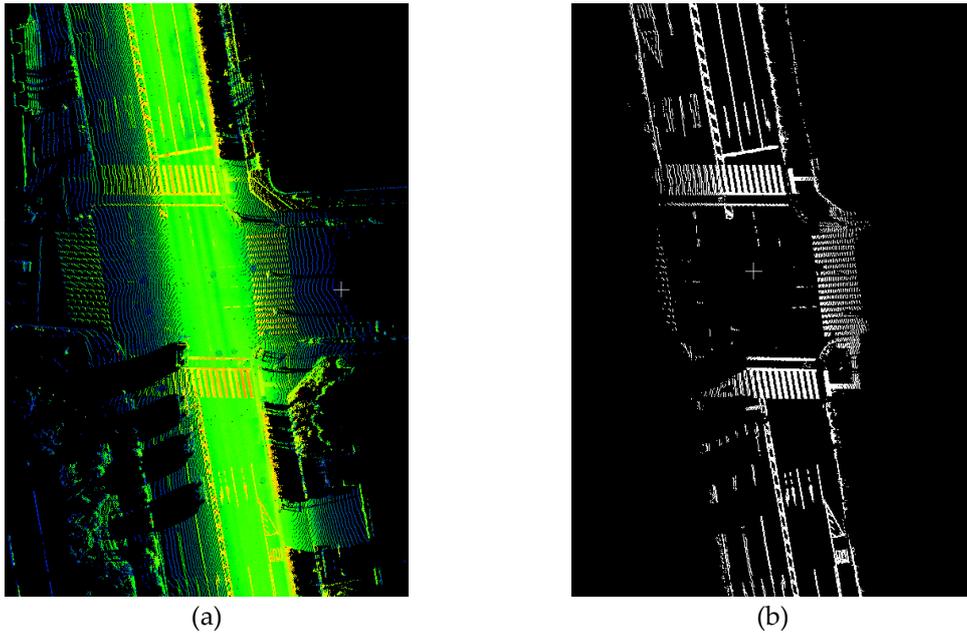
**Figure 3-11.** Intensity calibration of MMS point cloud: (a) original intensity of the MMS point cloud; (b) calibrated intensity using the proposed method.

### 3.2.3.3 Road marking extraction

After the calibration of the intensity, we should extract the similar road markings from the MMS point cloud for the registration step. MMS road markings are derived by applying the adaptive thresholding. **Figure 3-12** shows a sample road markings extracted from MMS point cloud, where the white points are the road markings.

### 3.2.4. 2D registration

Thus far, we have obtained the road markings from both aerial images and MMS surveys. The final step is to perform accurate and precise registration of each MMS survey to the aerial image. We have two different input formats. The first input is the aerial road markings in the form of a two-dimensional bitmap image, while the other input consists of MMS road markings in a three-dimensional point cloud format. To perform the registration, we proposed a method based on a dynamic-length sliding window and NDT. In this approach, a Gaussian mixture map is first generated from the aerial road markings



**Figure 3-12.** Road marking extraction from MMS point cloud: (a) original MMS point cloud (the color represents the original intensity); (b) the result of road marking extraction.

as a reference. Then, the sliding window is shifted over the patch-wise cropped MMS road markings to consecutively perform the registration of each MMS survey. Finally, the obtained registration result is employed to align the original MMS point cloud and image data.

#### 3.2.4.1 Generating an NDT Map from Aerial Road Markings (Reference)

The main idea behind the NDT is to represent the observed spatial information as a Gaussian mixture model [79]–[81]. To prepare the reference data for the registration, the bitmap image of the airborne road markings is divided into a uniform grid with a predefined cell size. Assuming that the set of all  $n$  points within the cell  $c_{ij}$ ,  $\{p_1, p_2, \dots, p_n \mid p_k = (x_k, y_k)\}$ , has been drawn from a normal distribution  $N(\mu, \Sigma)$ , the maximum-likelihood estimate of the mean ( $\mu$ ) is defined as follows:

$$\mu = \frac{1}{n} \sum_{i=1}^n p_i, \quad (5)$$

and the maximum-likelihood estimator of the covariance matrix ( $\Sigma$ ) is defined as

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^n (p_i - \mu)(p_i - \mu)^T, \quad (6)$$

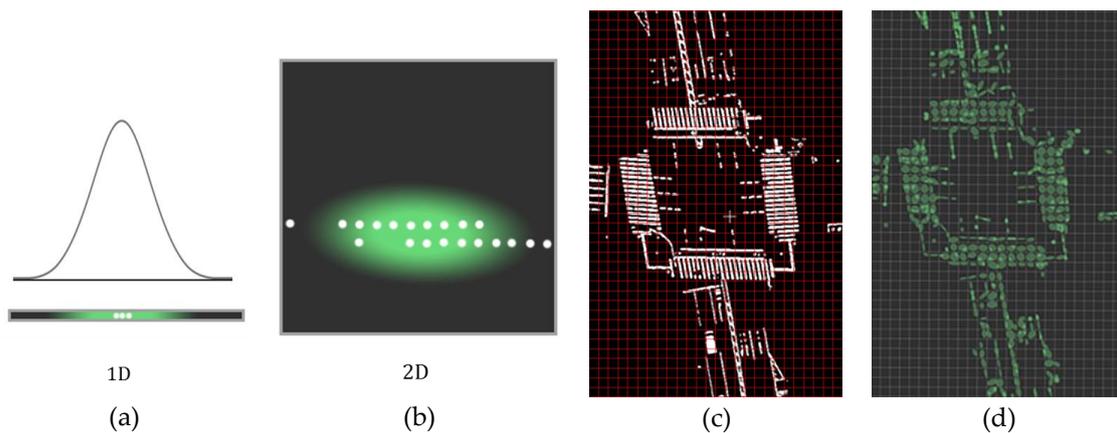
**Figure 3-13a** and **Figure 3-13b** illustrate the representation of the estimated Gaussian distributions for 1D and 2D cells. **Figure 3-13c** shows a part of the aerial road markings, and the generated Gaussian mixture map is shown in **Figure 3-13d**.

The probability of observing a road marking at a certain 2D coordinate of the cell  $c_{ij}$  is derived from the following equation:

$$P(p) \sim e^{-\frac{(p-\mu_{ij})^T \Sigma_{ij}^{-1} (p-\mu_{ij})}{2}}, \quad (7)$$

where  $p$  presents the 2D coordinate  $(x, y)$  of the point.

The size of NDT grid is an important parameter for the registration. If the grid size is set to be larger than the suitable dimension, the defined probability density function for the cell might not be able to represent well the points included in the cell. In contrast, if the grid size is small, each grid can capture only a few number of input road marking points, so there will be only a few grids having the normal distribution [82]. Therefore, a

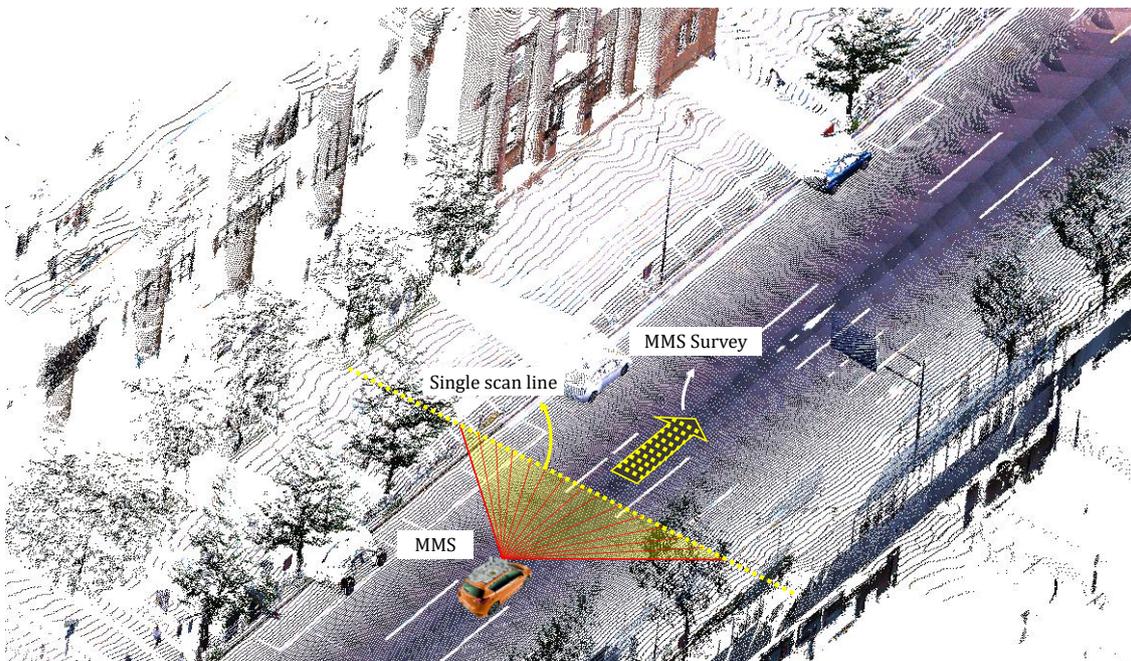


**Figure 3-13.** Generation of the Gaussian distribution from the reference data: (a) estimated 1D Gaussian distribution of the sample points; (b) estimated 2D Gaussian distribution of the sample points; (c) extracted aerial road markings; (d) generated NDT map with a 2 m grid size.

small error will make the input points out of the corresponding reference grid.

#### 3.2.4.2 Dynamic Sliding Window over MMS patches (Input)

The original MMS surveys are generated by a direct georeferencing technique that registers the Lidar measurements using the positions and orientations acquired from deeply coupled GNSS, IMU and odometer measurements [83]. Thus, each survey is made from thousands of scan lines where the relative positioning accuracy between two consecutive scan lines is at a centimeter-level. **Figure 3-14** illustrates the definitions of the laser scan line and MMS survey. The absolute accuracy of the survey is defined by the continuous quality of the scan line registration, which is affected by satellite visibility, the performance of the IMU, and driving conditions. In deep urban areas where tall buildings and other structures surrounding the streets, the GNSS/IMU positioning accuracy is significantly degraded. In this case, the overall accuracy of each MMS survey is mainly affected by the error budget of GNSS/IMU integration [84], which sometimes

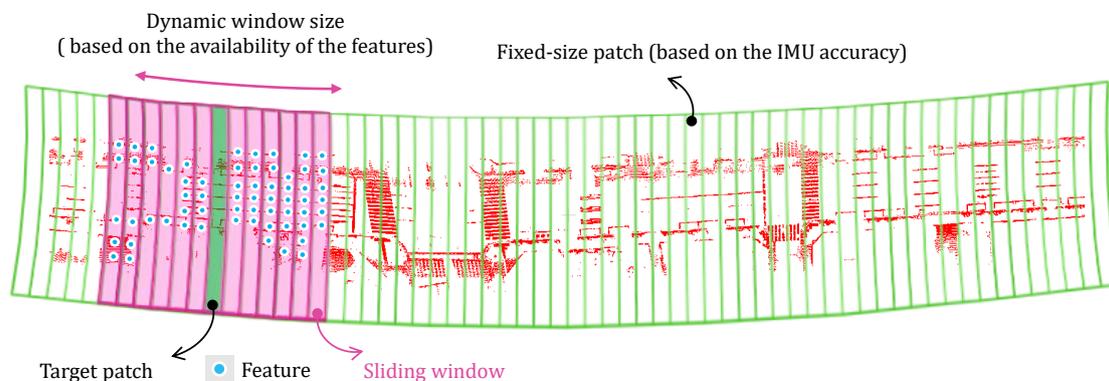


**Figure 3-14.** Illustration of single laser scan and MMS survey.

exceeds a few meters.

The relative error within each scan line is centimeter-level; therefore, the scan line can be considered as a rigid-body during the registration. However, due to the uncertainty in the bias of the IMU, a survey consisting of thousands of scan lines cannot be treated as a rigid-body. Ideally, to obtain a high-accuracy point cloud from the survey, every single scan line should be aligned to its correct absolute position based on the road features. However, a single scan line provides only a few recognizable features for the registration. To overcome this problem, rather than considering the scan line as the unit of registration, short and fixed-length patches consisting of several grouped scan lines are considered. The size of the patch should be defined based on the precisions of the IMU and the odometer to limit the relative error between the points within a patch (e.g., 1 cm relative error). To obtain the patches, the proposed method subdivides each MMS survey as shown in **Figure 3-15**.

These short pieces still do not contain sufficient road markings, especially to accomplish longitudinal registration. Thus, to define the position of the target patch, we consider a window that includes tens of patches around the target patch. This window should be sufficiently large to cover an adequate number of features for the registration



**Figure 3-15.** Concept of the dynamic sliding window.

process. The length of the window is an important parameter to achieve the required quality. If the length is too short, registration quality will suffer from a lack of road markings in the window. On the other hand, if a long window is adopted to cover more road markings, the relative accuracy of the scans inside the window will decrease due to the uncertain bias error of the IMU, which significantly degrades the quality of the matching. In this research, instead of defining a fixed-length window, we set the window length dynamically based on the availability of a sufficient number of road marking features within the window. The concepts of the patches, a target patch, the dynamic window and features are illustrated in **Figure 3-15**.

The window length is defined for each target patch based on three predefined factors: (1) patch length; (2) initial window length; (3) required feature count. Initially, the window is defined to include a predefined number of patches. Then, the length of the window is extended by adding new patches until it covers the required number of features. Finally, the window is registered to the aerial image to calculate the transformation matrix, which aligns the target patch. After each registration, the obtained transformation matrix is stored, and the window shifts to the next patch, where its length is redefined based on the new environment.

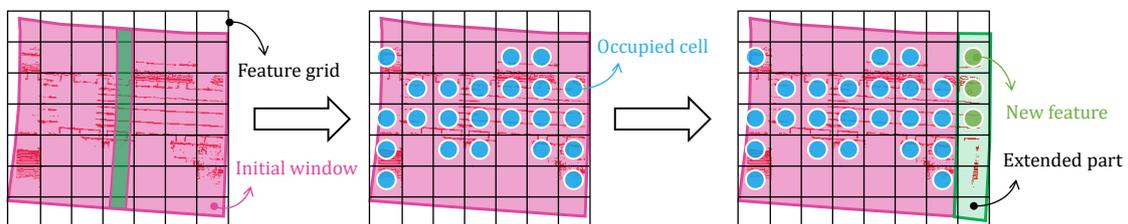
In contrast, the patch length is fixed and is considered as the unit of the registration. On one hand, choosing a long patch limits the achievable accuracy. On the other hand, the registration accuracy is limited by the aerial image's resolution. Therefore, even very small patches cannot improve the accuracy beyond a certain level, but very small patches increase the processing time. Thus, a relatively short patch length should be chosen to maintain high accuracy within a reasonable processing time. The initial window length defines the initial number of patches around the target patch, but the

required feature count defines whether the initial window should be extended. To count the existing features in a window, we subdivide the window into feature grids and calculate the number of occupied cells (five or more points). If the occupied cells exceed the required feature count, the window length is sufficient. If not, a new patch is added to the end of the window, and the number of occupied cells is recalculated. This cycle repeats until the window covers the required number of features. **Figure 3-16** illustrates these steps in sequence. Using the proposed method, the window length will increase in areas where only a few road markings are available. This enables a robust registration for such areas, in which other methods cannot find a reliable approach to conduct the transformation between the mobile mapping data and the aerial images. Moreover, at crossings where an abundance of road markings exists, the window is kept small to achieve highly accurate matching.

Because the window length is much longer than the sliding interval (which is the patch length), two consecutive scans have a large overlap. This overlap provides a smooth registration result.

### 3.2.4.3 2D MMS Registration

In this section, details of the MMS registration is described. To obtain the



**Figure 3-16.** The procedure of defining the window length for the target patch. First, the initial window around the target patch is divided into a feature grid (left). Second, the occupied cells—those containing five or more points—are calculated (middle). The window is extended by adding new patches until the number of occupied cells meets or exceeds the required feature count (right).

transformation matrix  $M_{t_i}$  (Equation 8) which is applied to the target patch  $t_i$  to define the position correction vector and generated the adjust patch  $t'_i$  (Equation 9), the target window  $W_{t_i} = \{\vec{w}_{t_{i1}}, \dots, \vec{w}_{t_{in}}\}$  should be registered to the aerial NDT map.

$$M_{t_i} = \begin{bmatrix} R_z(\theta) & T_{x,y} \\ 0 & 1 \end{bmatrix}, \quad (8)$$

which is applied to the target patch  $t_i$  to generate the aligned patch,

$$t'_i = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} t_i + \begin{bmatrix} T_x \\ T_y \end{bmatrix}, \quad (9)$$

the window  $W_{t_i} = \{\vec{w}_{t_{i1}}, \dots, \vec{w}_{t_{in}}\}$  should be registered to the aerial map.

To register the window  $W_{t_i}$ , the points inside it are first subsampled and then transformed to absolute coordinates based on an initial guess. While the initial guess for the first window is obtained through GNSS/IMU estimation, the transformation matrix of the previous window  $M_{t_{i-1}}$  is applied to the GNSS/IMU position of the current target patch  $t_i$  to form the initial guess for the  $W_{t_i}$ . To estimate the matching score, a Gaussian mixture model is used instead of a normal distribution to limit the effect of outliers, which significantly decrease the output of the normal distribution [80]. Therefore, instead of Equation 7, a mixture of uniform and normal distribution is used:

$$\hat{P}(\vec{x}) = c_1 P(\vec{x}) + c_2 P_0, \quad (10)$$

where  $P_0$  is the expected rate of the outliers, and the constants  $c_1$  and  $c_2$  are defined to normalize the distribution. Finally, the score of the point  $\vec{x}_k$  is calculated by applying it to the log-likelihood function and approximating it to derive the simplicity:

$$S(\vec{x}_k) = d_1 \exp\left(-d_2 \frac{(\vec{x}_k - \mu_j)^T \Sigma_j^{-1} (\vec{x}_k - \mu_j)}{2}\right), \quad (11)$$

where  $j$  is the index of the closest Gaussian distribution, and  $d_1$  and  $d_2$  are obtained from  $c_1$  and  $c_2$ . Finally, the cost function of the window  $W'_{t_i} = \{\vec{w}'_{t_{i1}}, \dots, \vec{w}'_{t_{in}}\}$ , the result of registering  $W_{t_i}$ , is defined as follows:

$$C(W'_{t_i}) = -\sum_{k=1}^n \vec{w}'_{t_{ik}}. \quad (12)$$

Moreover, to optimize the cost function and achieve the best matching result, the Newton optimization method is employed [80]. After obtaining the transformation matrix from the road-marking registration, the original MMS data of the corresponding patch are calibrated based on the same matrix.

### 3.3. Experiment and analyses

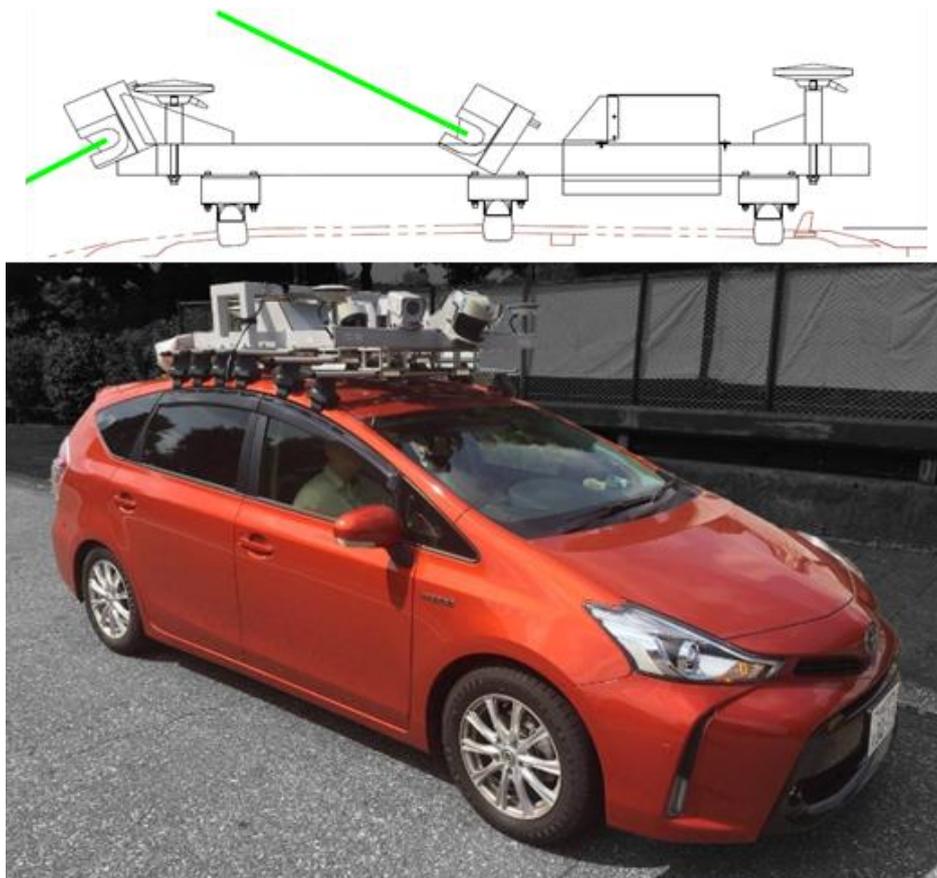
This section covers the experimental results of the proposed framework. The experimental setup, experimental area, and analyses are described in detail to evaluate the approach.

#### 3.3.1. MMS, aerial surveillance system

In this section, details of the MMS and aerial surveillance system, two main data acquisition platforms used for the collection data, are described.

##### 3.3.1.1 Mobile mapping system

For the MMS data collection of the experiments, we used an MMS-K320 system developed by Mitsubishi Electric. As shown in **Figure 3-17**, this system is equipped with two single-layer laser scanners and three cameras to perform a 3D measurement of the surroundings, two single frequency GPS receivers, one dual frequency GPS receiver, a high-end IMU, and an odometer for a precise vehicle self-localization. Two laser scanners with the field of view of  $180^\circ$  were configured to see up (pitch:  $+25^\circ$ ) and down (pitch:  $-25^\circ$ ). The laser scanners could collect reflective luminance which is a measure of the return strength of the laser pulse that generated the point. The full specification of the



**Figure 3-17.** MMS system description: Mitsubishi Electric’s MMS-K320 (bottom) and the configuration of two SICK LMS-511 laser scanners and RTK GPS receivers (top).

system can be found on the Mitsubishi Electrics website.

### 3.3.1.2 Aerial surveillance system

Our aerial surveillance system was equipped with a medium format single head camera (Leica RCD30 series 80 MP multispectral imagery) installed in a direct down-facing position to acquire the high-resolution imagery, a Lidar system (Leica ALS70-HP) with a 500 kHz pulse rate for 3D urban mapping, and a high-end GNSS/IMU for positioning. The aircraft altitude in the experiments was more than 1700 m, the ground sampling distance (GSD) of the acquired aerial images was equal to 12 cm, and the point density of the obtained ALS point cloud was less than 10 pts/m<sup>2</sup>. **Figure 3-18** shows the aerial system. **Table 3-1.** Details of the sensor platforms employed for the experiments

lists the details of the sensor platforms.



(a)



(b)

**Figure 3-18.** Aerial surveillance system description: (a) the aircraft used for aerial data collection; (b) the sensor setup for the aerial imagery.

**Table 3-1.** Details of the sensor platforms employed for the experiments

Item		Value	
MMS*	Laser Scanner	Manufacturer (Model)	SICK (LMS-511)
		No. mounted	2 single layer lasers
		Mounting direction	CH1: Front/Down (-25°), CH2: Front/Up (25°)
		Intensity	Can be acquired
		No. of points	27,100 points/sec (1 unit)
		Range (max.)	65 m
		Viewing angle	180°
	Camera	No. mounted	3
		No. of pixels	5 megapixels
		Max capture rate	10 images/sec
		View angle	Wide viewing angle (h: 80°, v: 64°)
	Localization platform	Manufacturer	Mitsubishi Electronics
		Method	RTK-GPS/IMU/odometer
	Self-positioning accuracy *1		Within 6 cm (rms) *4, 5
Relative accuracy data *2		Within 1 cm (RMS)	
Absolute accuracy data *1, 3		Within 10 cm (RMS)	
Flying height		~1700 m	
Aerial system	Laser Scanner	Manufacturer (Model)	Leica (ALS70)
		Mounting direction	Direct-down
		Intensity	Can be acquired
		Max measurement rate	500 kHz
		Point cloud density *6	Less than 10 pts/m <sup>2</sup>
		Field of view	~75°
	Camera	Manufacturer (Model)	Leica (RCD30)
		Mounting direction	Direct-down
		No. of pixels	80 MP (10320 x 7752 pixels)
		Maximum frame rate	0.8 fps
		Forward overlap	60%
		GSD *7	12 cm/pixel

Localization platform	Manufacturer (Model)	Novatel (IMU-LN200)
	Method	GNSS/IMU

\* [www.mitsubishielectric.co.jp/pas/mms](http://www.mitsubishielectric.co.jp/pas/mms)

\*1 Assuming favorable GPS reception (RMS: root mean square).

\*2 Relative accuracy: the consistency of the coordinate values captured during mobile measurements.

\*3 Absolute accuracy: the extent to which the coordinate values captured during mobile measurements match actual coordinate values.

\*4 Driving on a level road at a constant speed of approx. 40 km/h.

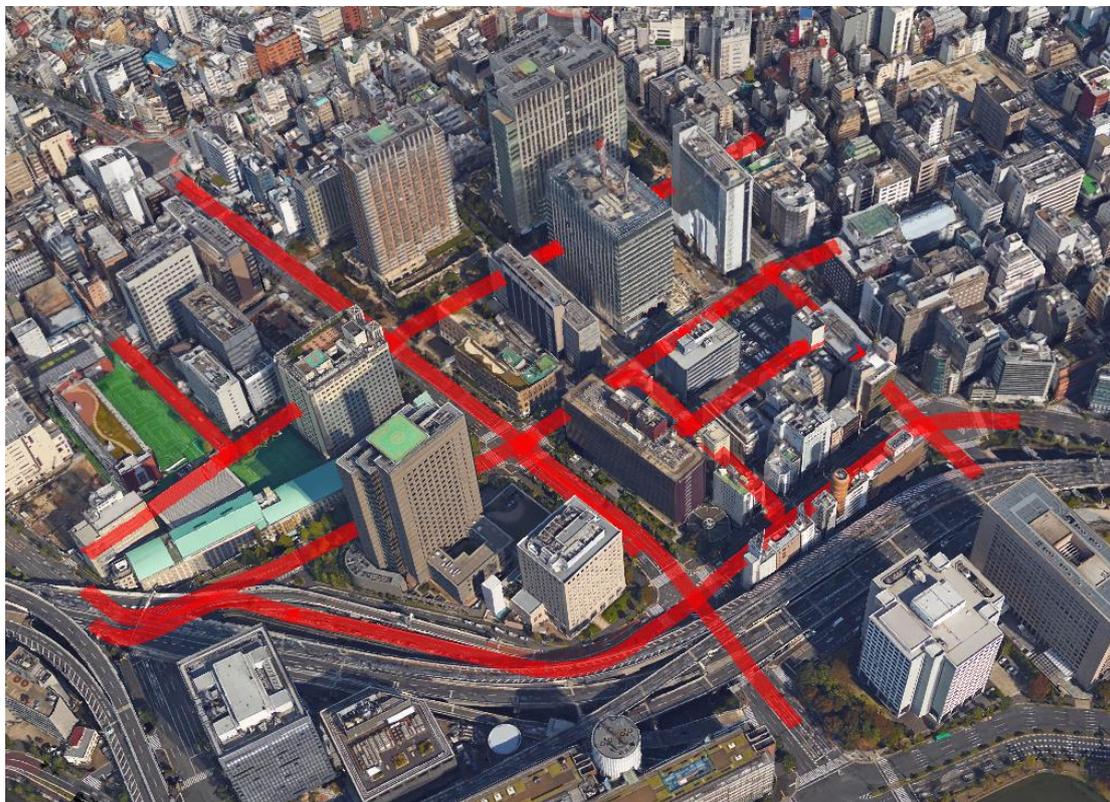
\*5 User must calibrate before each survey.

\*6 Point density is defined by the flying height of the aircraft and the scanner measurement rate.

\*7 GSD is defined by the flying height of the aircraft and the camera resolution.

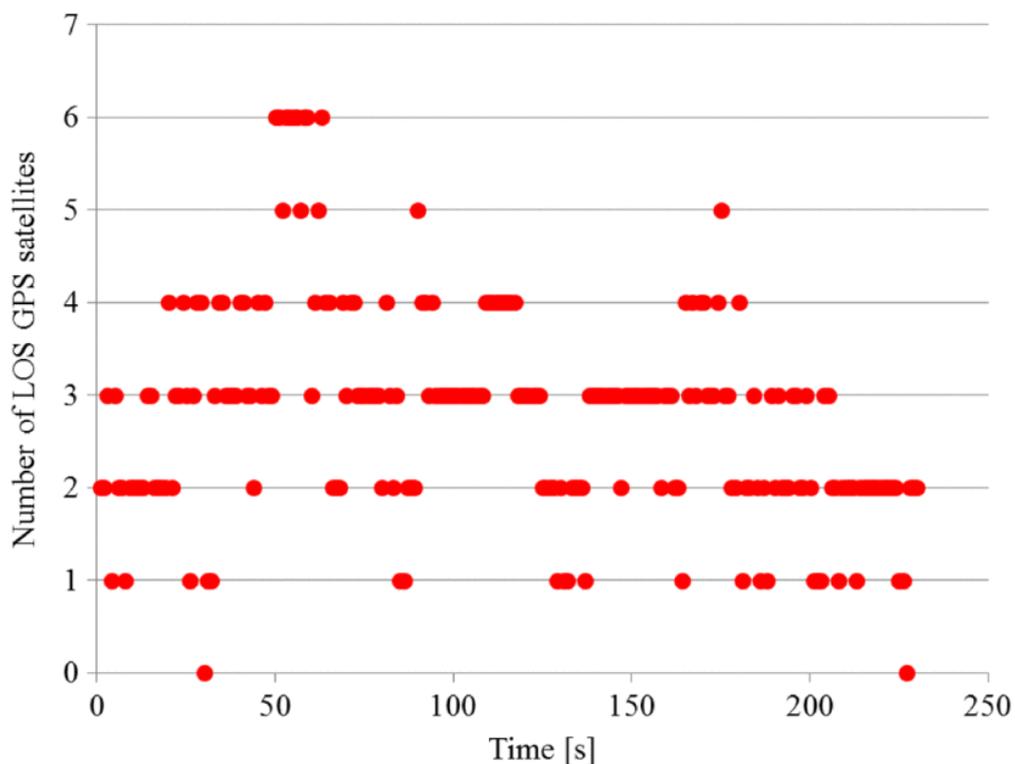
### 3.3.2. Experimental area

To evaluate the performance of the proposed system, the MMS and aerial data were acquired in the vicinity of Hitotsubashi, a dense urban area in the Chiyoda-ku area



**Figure 3-19.** The experimental area around Hitotsubashi intersection, a dense urban area of the Chiyoda-ku, Tokyo, Japan.

of central Tokyo, Japan (**Figure 3-19**). Streets around the Hitotsubashi intersection are surrounded by tall buildings, trees, and traffic. A survey on the number of Line-of-Sight (LOS) GPS satellites in this area showed that two-thirds of the time, less than four satellites were visible (**Figure 3-20**) [85]. This shortage increases the positioning error to a few meters, which is unsuitable for precise mapping. Nine MMS surveys were performed to obtain ground data of the area; the survey routes included multiple crossings, traffic flow, and road markings on the pavement. The aerial survey was performed on June 12, 2014 (a cloudy day in spring), and the MMS data were acquired on April 20, 2016 (a sunny day in spring). If the road markings are substantially altered or repainted to a different position after obtaining the aerial image, the registration performance will be affected. In our data, a few parts of the road markings were repainted after Jun 12, 2014 (the date of the aerial image acquisition). To simplify error interpretation, we



**Figure 3-20.** An example of number of LOS GPS satellites observed in the Hitotsubashi intersection in Tokyo, Japan.

manually updated the aerial road markings of those areas based on the newer observations, and report the effect of these road-marking changes on the performance of the framework separately.

### 3.3.3. Experimental parameters

This section provides an overview of the experimental parameters used for the final evaluation. **Table 3-2.** Summary of the parameters applied in our experiment lists the parameters.

**Table 3-2.** Summary of the parameters applied in our experiment

Parameters		Value	Description
POM Generation	Height threshold	5 m (above the ground)	Defined based on the maximum height of the vehicles
	Resolution	12 cm/pixel	Equal to the GSD of the aerial image
MMS ground segmentation	Cloth resolution	2 m	Larger grids do not cover the ground well (set empirically)
	Max iteration	1000	More than 500 is suggested
	Classification threshold	20 cm	If the cloth resolution is set correctly, small values give suitable results
Adaptive thresholding	Block size	2.5 m (21 pixels)	Empirically defined
	Threshold	weighted mean - c	c = 17 empirically defined
Dynamic sliding window	Patch length	0.5 m	Defined based on IMU performance to limit the error within 1 cm
	Window length	Dynamic	-
	Initial window length	60 patches	Equal to 30 m (60 × 0.5 m)
	Feature grid size	1 m	Empirically defined
	Required feature count	400	Empirically defined
NDT registration	NDT grid size	1 m	Defined to be smaller than the distance between the lane

			markings and signs in the middle of the lanes
	NDT iterations	30	Should be high enough to let the NDT converge (set empirically)

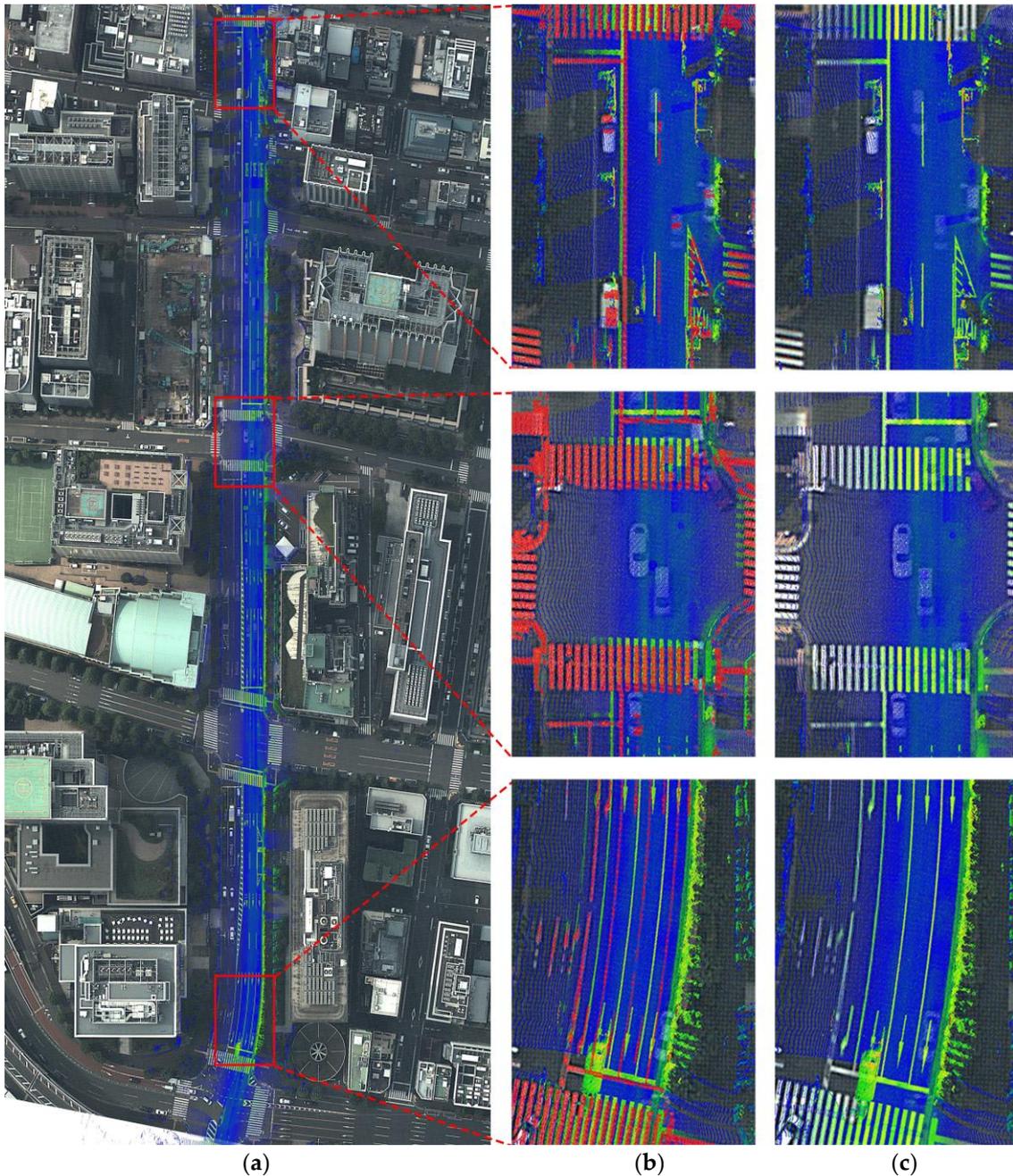
### 3.3.4. Experimental results

Nine MMS surveys were georeferenced using the proposed framework. The survey routes are shown in **Figure 3-21**. All the computations were conducted on an off-the-shelf PC with an 8-core, 3.50 GHz Intel Xeon E3-1270 V2 CPU and 16 GB of RAM running the 64-bit version of the Ubuntu 16.04 operating system. The average time required to apply the framework was 0.94 seconds per meter per survey, which included the registration, visualization, geo-referencing of the original data, and saving the updated data. We successfully calibrated all nine surveys using the proposed framework.



**Figure 3-21.** Routes of the nine surveys S01–S07 on the map. The colors show the estimated GPS error reported by MMS.

**Figure 3-22** illustrates the performance of the proposed method for survey No. 6 visually. The survey's length was approximately 557 m. In **Figure 3-22**, the color of the MMS point cloud represents the intensity of the lidar reflection. In **Figure 3-22b**, the



**Figure 3-22.** Visual evaluation of the proposed method for survey No. 6: (a) survey route of the MMS on the aerial image (after registration); (b) enlarged view of the original MMS point cloud before registration (the red points are aerial road markings); (c) enlarged view of the MMS point cloud after applying the proposed method.

aerial road markings are highlighted in red to show the shift between the aerial image and MMS data before registration. The MMS data and aerial image clearly overlap (Figure 3-22c).

### 3.3.5. Discussion

The MMS-K320 is able to report the estimated localization error during the experiment to allow the operator to increase the accuracy of data collection. Usually, when the operators are notified that the localization error exceeds 2.5 meters, they should suspend the survey and move the vehicle to a location with sufficient satellite signals and wait for the convergence. Then, the survey must be restarted from the beginning. However, in dense urban areas like Hitotsubashi intersection, in most cases, even after a move to an open-sky area to cause GPS convergence, the error would exceed 2.5 meters before the vehicle arrived back at the start point of the survey. Figure 3-23 shows the estimated GPS error for each of the nine surveys.

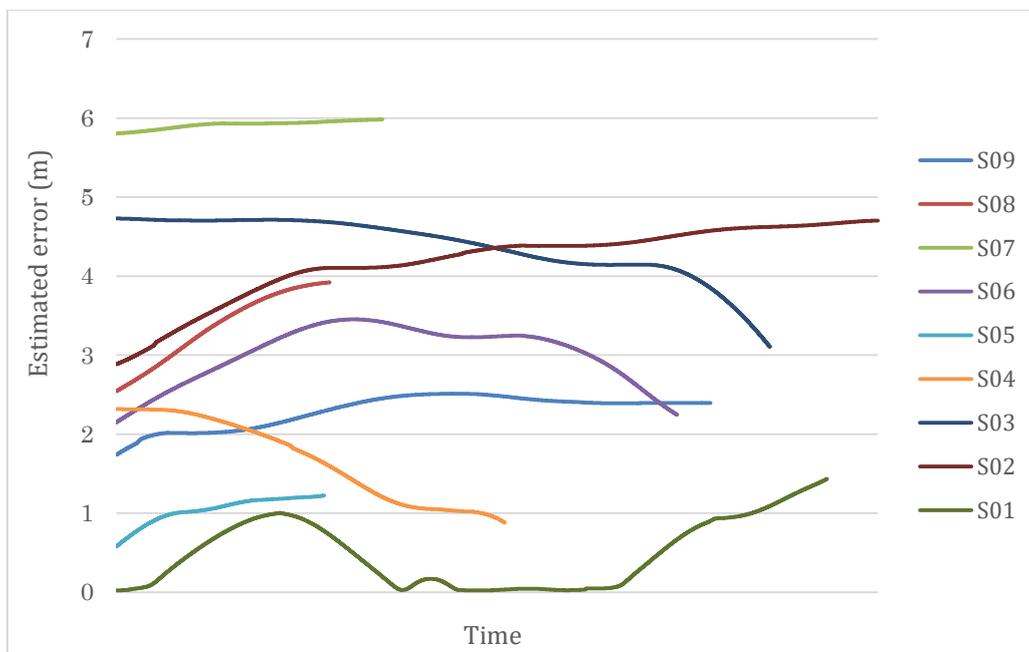
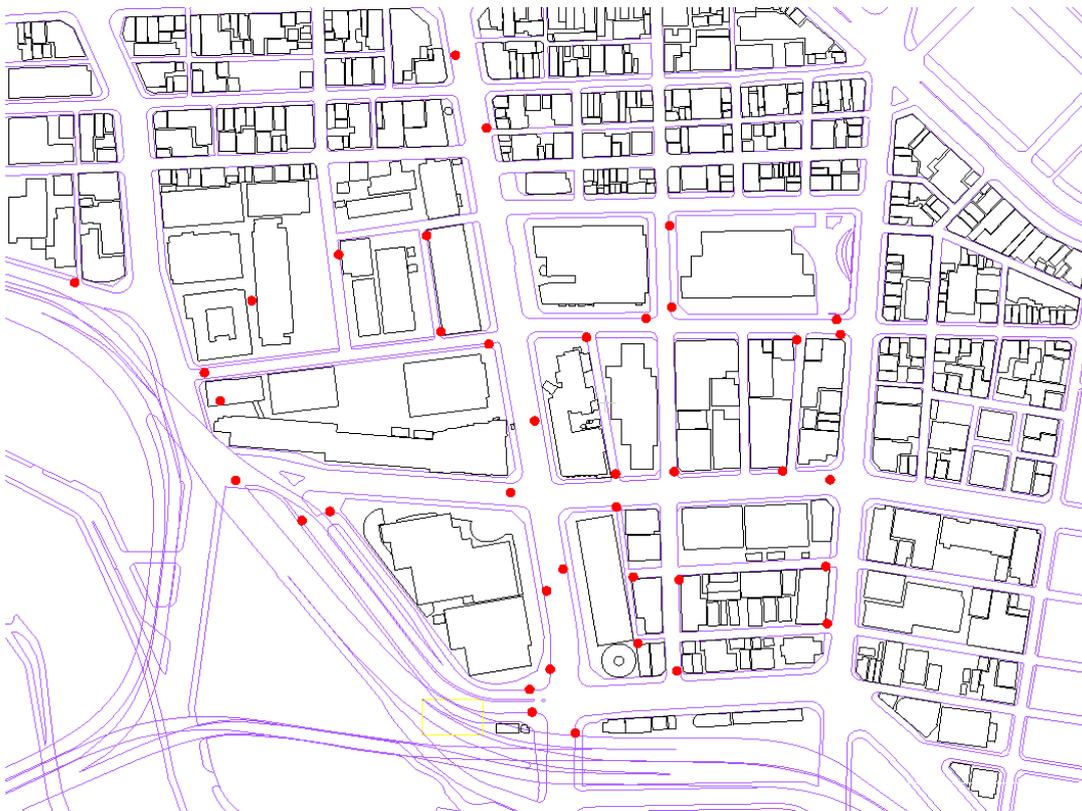


Figure 3-23. Estimated GPS error by MMS-K320 for each of the nine surveys.

To evaluate the performance of the proposed framework, 39 GCPs at the corners of road signs were measured using the total station surveys in the area, and 136 virtual control points (VCPs) were generated from the original GCPs and the aerial image. **Figure 3-24** illustrates the GCP distribution in the test field. Additionally, a labor-intensive landmark updating procedure was performed to allow a comparison of the result with the proposed MMS geo-referencing framework. For this purpose, the original GCPs were manually picked from the MMS point cloud using software provided by the company to obtain the calibrated point cloud. The 2D error  $d$  was calculated for all VCPs using the following equation:

$$d_i = \sqrt{(X_{VCP_i} - x_i)^2 + (Y_{VCP_i} - y_i)^2}, \quad (12)$$



**Figure 3-24.** Distribution of the GCPs collected in the experimental area utilizing the total station survey. The GCPs are from the corner points of road signs, which are clearly captured by both the MMS camera and the point cloud.

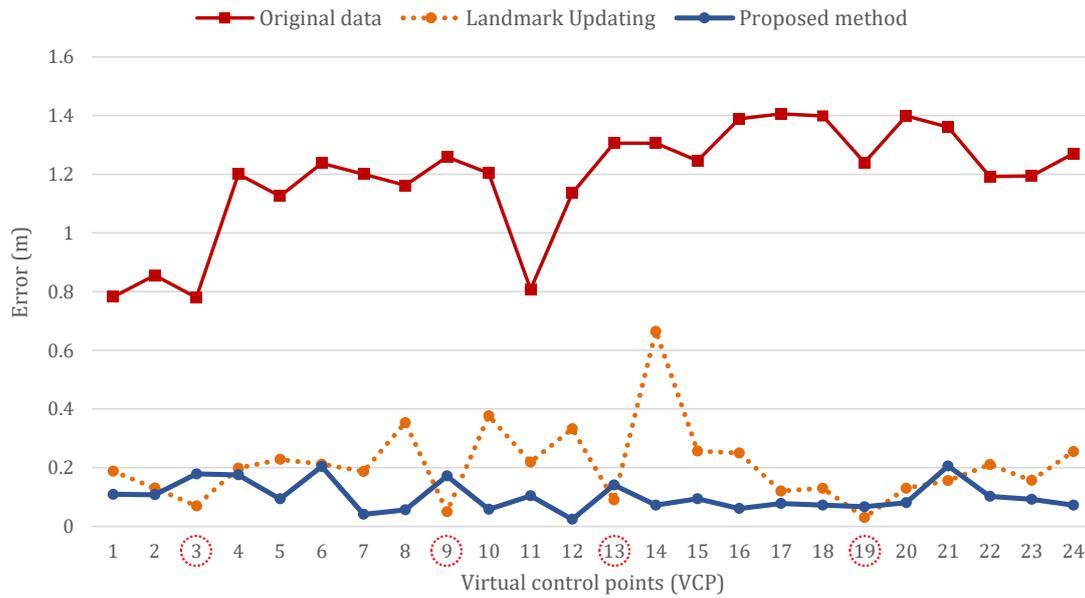
where  $(X_{VCP_i}, Y_{VCP_i})$  is the 2D coordinate of the  $i^{th}$  VCP, and  $(x_i, y_i)$  is the coordinate of the point corresponding to the  $i^{th}$  VCP in the MMS point cloud. For the quantitative evaluation, the mean, maximum and standard deviation of the calculated 2D error are reported for the original MMS data, landmark updating and the proposed approach.

**Figure 3-25** shows a comparison of the proposed method, landmark updating and the original MMS data for survey No. 6, which extended for 557 meters from the north to the south of the Hitotsubashi intersection. The proposed method successfully limited the average error to below 12 cm (the resolution of the aerial image), while its maximum error was 20.6 cm. The average error of the landmark updating method was 20.8 cm, while its maximum error exceeded 66 cm. The average error in the original data was reported as 118.6 cm, and the maximum error exceeded 140 cm. As shown, the proposed method outperformed the labor- and time-intensive landmark update, due to the availability of sufficient road features and their relatively uniform distribution along the survey route. **Table 3-3** summarizes the compared results.

In **Figure 3-25**, the landmark updating method resulted in higher accuracy than our method for a few virtual control points, which are highlighted by red circles. In fact, the coordinates of those VCPs are close to the original GCPs used for landmark

**Table 3-3.** Evaluation result of survey No. 6

Method	Error (m)		
	Mean	Max	Stdev
Original data (GPS/IMU/Odometer)	1.186	1.405	0.19
Landmark updating	0.208	0.664	0.11
Proposed method	0.102	0.206	0.05



**Figure 3-25.** The 2D error of survey No. 6 in the original data, after the landmark update, and after the proposed method

updating. Because the objective of the landmark updating technique is to update the vehicle’s trajectory to minimize the error in the incorporated GCPs, it performs very well for areas around the GCPs. Therefore, its good performance was not unexpected. However, the proposed method yielded a better performance in most cases.

The proposed framework depends on several variables and parameters. In particular, the final registration step is sensitive to a variety of parameters such as the NDT grid size, the initial window length, the required feature count, and the patch length. We evaluated the sensitivity of the proposed framework to all four of these parameters. To simplify the interpretation of the error behavior, we evaluated each parameter by assigning different values to it while using the default values for the other parameters.

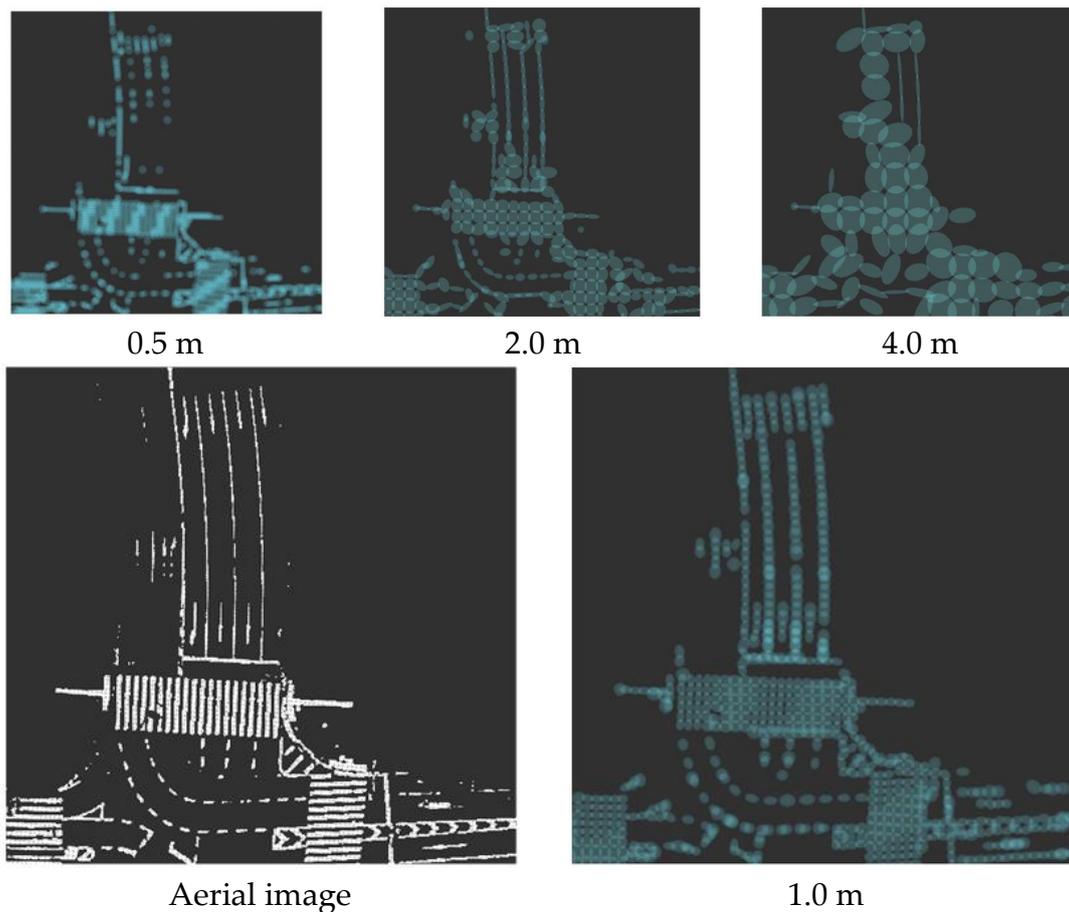
**Table 3-4.** The proposed framework's sensitivity to different NDT grid sizes shows the registration result of survey No. 6 for various NDT grid sizes. As shown, grid size plays an important role in registration quality.

On one hand, when the grid size is too small (e.g., 0.5 cm), some grids will

**Table 3-4.** The proposed framework's sensitivity to different NDT grid sizes

Item	Value			
	Proposed			
NDT grid size* (m)	0.5	1	2	4
Mean error (m)	5.01	0.11	1.20	1.22

contain insufficient points to calculate the Gaussian distribution. Therefore, some important road marking information will not be used in the registration. On the other hand, if large grids are selected (e.g., 2.0 or 4.0 m), a Gaussian distribution will not provide a good estimate of the local environment (which is road markings). **Figure 3-26** shows the Gaussian mixture models generated using different grid sizes.



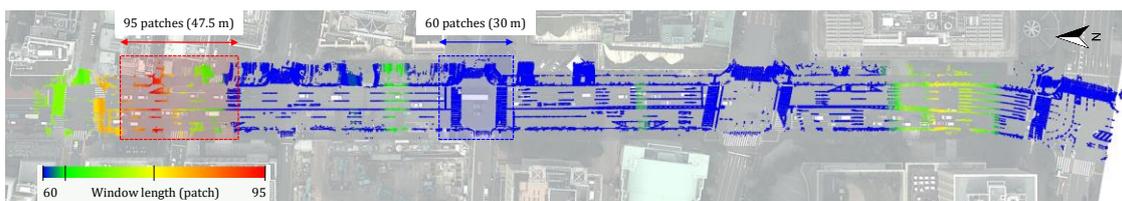
**Figure 3-26.** Generated Gaussian mixture models using different grid sizes.

**Table 3-5.** Evaluation of the sensitivity of the framework to different initial window lengths

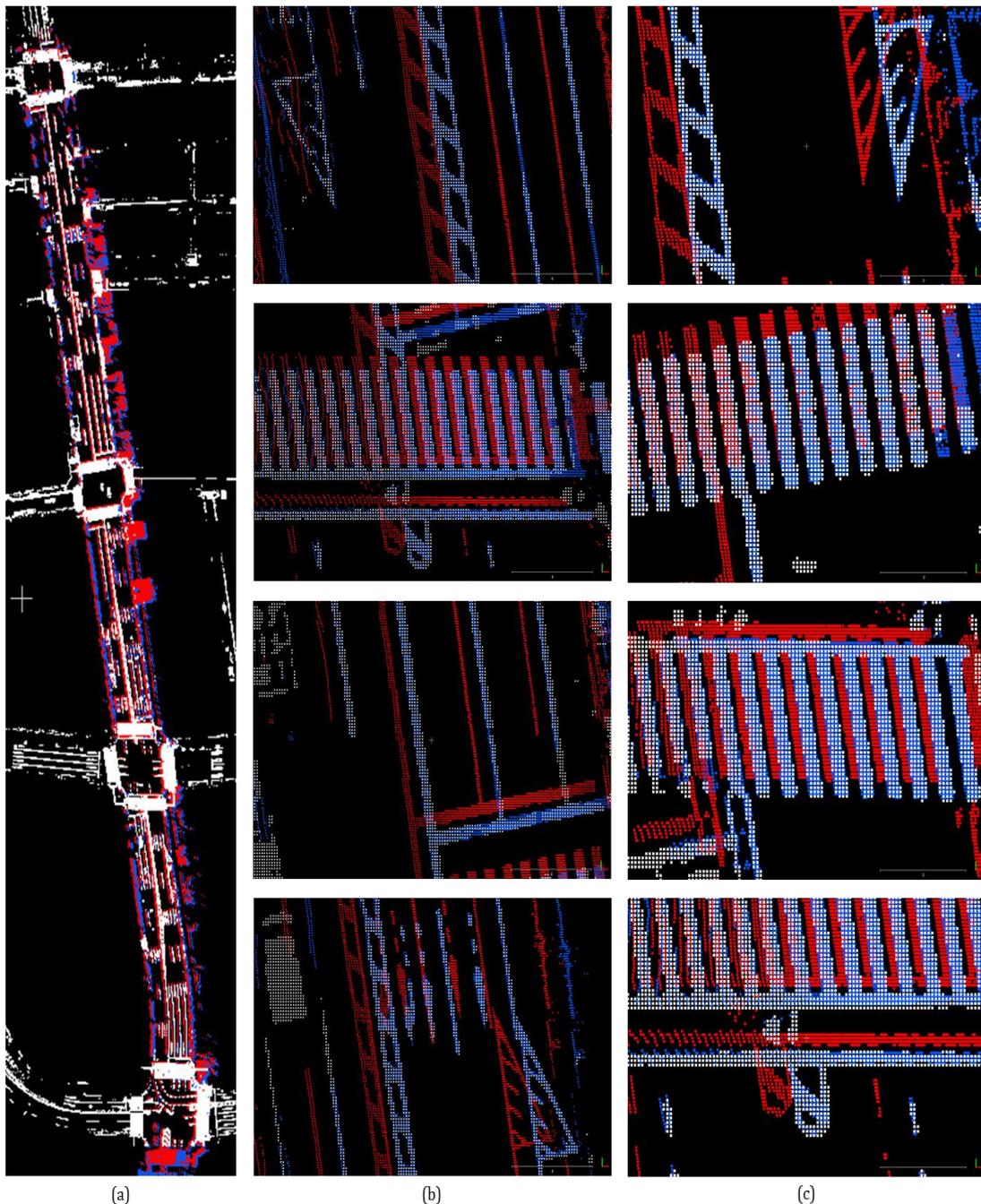
Item	Value			
	15	30	Proposed 60	120
Initial window length* (patch)	15	30	60	120
Min win length (patch)	39	39	60	120
Mean win length (patch)	59.22	59.16	64.67	120
Max win length (patch)	95	95	95	120
Mean error (m)	0.80	0.52	0.11	3.34

The window length is another parameter that affects the registration performance.

**Figure 3-27.** illustrates how the length of the dynamic sliding window changed for survey No. 6. **Table 3-5** and **Table 3-6** show how the initial window length and required feature count affect the window length and the registration error, respectively. A short initial window length (e.g., 15 and 30 patches) cause the registration to become sensitive to false positive features and decrease the matching accuracy. On the other hand, a large initial window (e.g., 120 patches) with a fixed required feature count will result in a long and static window length, which also affects the matching significantly. In the proposed framework, the initial window length was set to 60 patches, which provides the dynamic windows with a length of at least 30 m. After selecting a suitable initial window length, the required feature count should be defined so that it allows extension of the initial window in the absence of road markings. If the required feature count is too small (e.g., 200), the window length will remain equal to the initial window length. On the other hand,

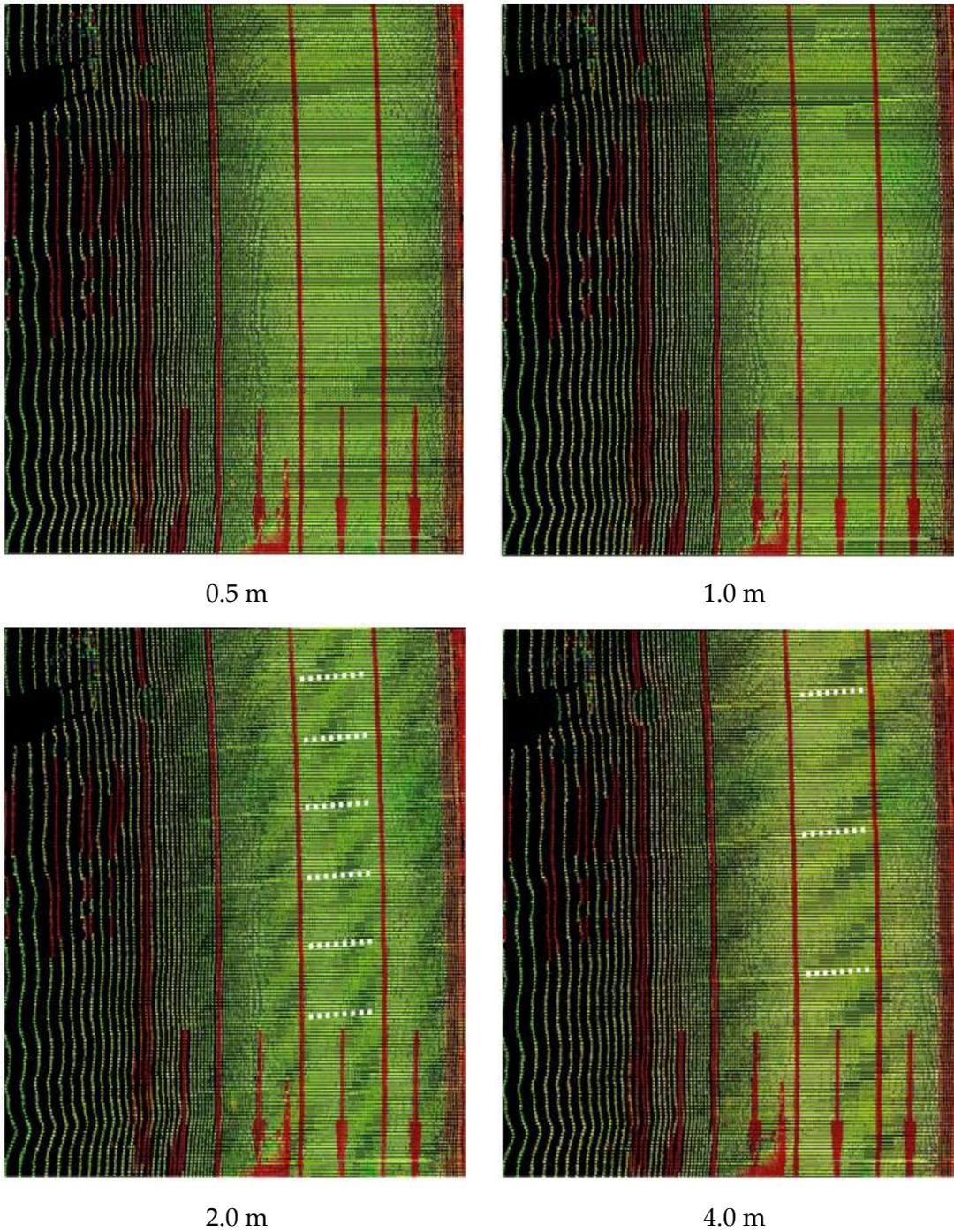
**Figure 3-27.** Length of the dynamic sliding window for survey No. 6

large feature counts (e.g., 600 or 800) will extend the window length disproportionately, which decreases the matching accuracy (Table 3-6). Figure 3-28 shows another visual evaluation of proposed method.



**Figure 3-28.** Visual evaluation of the proposed method: (a) survey route of the MMS on the aerial image; (b) enlarged view of the original (red) and calibrated (blue) MMS point cloud on the aerial image (the scale is 4 m); (c) enlarged view of the original (red) and calibrated (blue) MMS point cloud on the aerial image (the scale is 3 m).

Finally, the patch length is important to obtain a smooth point cloud from the registration. **Figure 3-29** illustrates the georeferenced MMS point cloud using different



**Figure 3-29.** Effect of the patch length on the smoothness of the georeferenced MMS point cloud

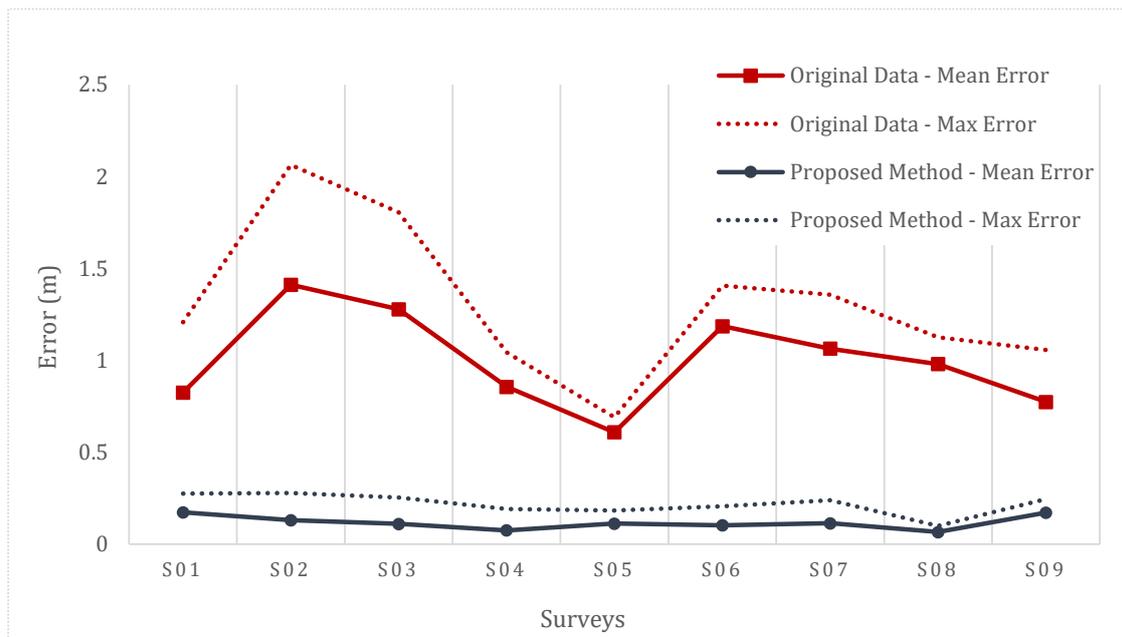
**Table 3-6.** Evaluation of the sensitivity of the framework to different numbers of required features

Item	Value			
	Proposed			
Required feature count*	200	400	600	800
Min win length (patch)	60	60	69	90
Mean win length (patch)	60	64.67	90.13	119.68
Max win length (patch)	60	95	130	165
Mean error (m)	2.37	0.11	0.72	3.58

\* All parameters other than the required feature count were chosen based on **Table 3-2**.

patch lengths. As shown, long patches (e.g., 2 and 4 m) result in a non-smooth point cloud. Although a smaller patch length is preferred, smaller patch lengths linearly increase the execution time.

**Figure 3-30** illustrates the overall performance of the proposed method for all nine surveys. In total, the proposed method decreased the average MMS data error from 99.7 cm to 11.6 cm while limiting the maximum error to 27 cm. The maximum error

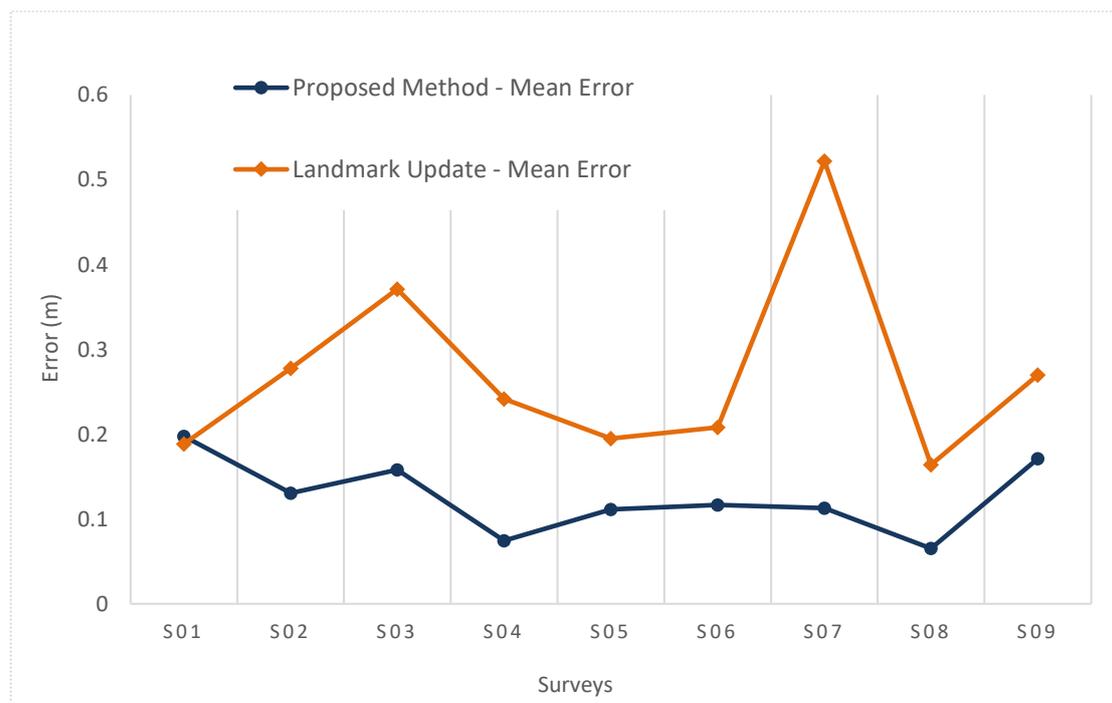


**Figure 3-30.** Total 2D error of the nine surveys in the original data and by the proposed method

reported for the original data was 206 cm.

**Figure 3-31** compares the overall performance of the proposed method with the landmark update method for all nine surveys. As shown, the proposed method not only significantly outperforms the traditional landmark updating method but also does not require the labor-intensive and time-consuming work involved in the conducting the GCP survey and manually finding the corresponding points in the data.

**Table 3-7** summarizes the overall performance of the proposed method in comparison with other state-of-the-art methods. The main advantage of the proposed method compared to the work presented in Kümmerle et al. [62] is its accuracy. In contrast, although we could achieve a higher accuracy than Hussnain et al. [65], our method's main advantage is that the proposed framework is robust to missing parts of road markings. Because Hussnain et al. used a feature-based registration technique with fixed size windows (tiles), their method is susceptible to insufficient features during registration.



**Figure 3-31.** Total 2D error of the nine surveys by the landmark update method and by the proposed method

Table 3-7. Overall performance comparison

Method	Error (m)		
	Mean	Max	Stdev
Original data (GPS/IMU/Odometer)	0.997	2.064	0.22
Landmark updating	0.208	0.72	0.16
Proposed method	0.116	0.277	0.07
Original data for [35] (Graph SLAM)	1.3*	1.93*	-
Kümmerle et al. [35]	0.85*	1.47*	-
Original data for [38] (GNSS/IMU/Odometer)	2.13**	2.40**	0.13**
Hussnain et al. [38]	0.18**	0.32**	-

\*The reported numbers are based on reports from the text and Fig. 13 of the cited paper.

\*\*The reported numbers are extracted from Table 3-2 and Figure 18 of the cited paper. Because they reported only the accurate and least accurate results using manually selected corner points in their aerial images, we estimated the mean and max values using only the graph of those tiles.

We overcame this challenge by applying dynamic overlapping sliding windows for registration. The same group reported in Jende et al. [86] that they were able to obtain a reliable transformation for fewer than 23% of their image tiles. In contrast, our framework achieved an accurate registration for all nine surveys.

The proposed method's registration is based on the availability of sufficient road markings in both MMS and aerial surveys. Therefore, it cannot be performed under certain conditions, e.g., under bridges or on streets without road markings. Additionally, significant changes to road markings after the aerial survey, because of reconstruction or repainting to different positions, will affect the quality of the registration. However, the proposed method is robust to small changes. **Figure 3-32** illustrates the areas of this study where the road markings were repainted after the aerial image acquisition. To assess the effect of these road-marking changes, we also performed survey registration using the original aerial image (which included outdated road markings rather than the updated versions). In that case, the maximum error increased from 27 cm to 52 cm, but the average



**Figure 3-32.** Areas where the road markings were repainted after June 12, 2014 (the date of the aerial image acquisition).

error increased by only 1 cm. In recent years, aerial surveys have become a popular method of acquiring data for different applications. Therefore, different companies have begun providing regularly updated aerial data of cities, making the proposed framework a promising scheme for accurate urban mapping.

### 3.4. Conclusions

In this chapter, we have presented a novel and thorough framework for automatic georeferencing of mobile mapping system (MMS) data that is specially designed for urban areas. First, road markings are extracted from high-resolution aerial images based on a perspective occlusion map (POM) and adaptive thresholding. The POM is generated

by perspective projection of the digital surface model (DSM) obtained from airborne laser scanning (ALS) data. Moving vehicles, which could be mistakenly extracted as road markings, are filtered by comparing overlapping aerial photographs. Then, similar road markings are extracted from an MMS point cloud to perform registration by road segmentation, laser intensity calibration and adaptive thresholding over intensity value. Finally, the MMS geo-referencing was achieved by performing a precise registration of each MMS survey to the aerial reference, utilizing a dynamic overlapping sliding window and a normal distribution transform (NDT). In urban areas, the quality of global navigation satellite system (GNSS) measurements is significantly degraded due to blockages and reflection of the satellite signals. Therefore, integrated GNSS/IMU positioning has a significant error with considerable uncertainty. In such cases, a single MMS survey cannot be treated as a rigid body for the registration. The proposed dynamic sliding window-based registration overcomes this problem by keeping the window length as short as possible based on the availability of road markings in the target area. The experimental results from several surveys in a dense urban area in Tokyo were employed to assess the performance of the proposed method. In total, the proposed method could decrease the average MMS data error from 99.7 cm to 11.6 cm, while it reduced the maximum error from more than two meters to 27 cm. These results indicate that the proposed method's performance is even better than that of the labor-intensive and time-consuming landmark update process, yet the proposed method does not require a field survey to acquiring the ground control points (GCPs) and avoids having to manually find those points in the point cloud data.

The current proposed framework performs a two-dimensional registration of the MMS data. To achieve a high-definition 3D urban mapping, three-dimensional

registration is desirable for some applications. In the next Chapter, we will consider a three-dimensional calibration of the MMS data based on the fusion of aerial images and ALS point clouds.

## Chapter 4.

# Three-Dimensional Registration of Mobile Mapping System and Aerial Surveillance Data

In the 2D registration method, the MMS point cloud was georeferenced based on the registration of the MMS road markings to the road markings extracted from the aerial image. This method can improve the 2D accuracy of the MMS point cloud which is enough for 2D applications. However, the error in the z-axis is also significant and should be considered for the geo-referencing. In order to update the Z-axis of the MMS data, we require a 3D reference for the matching, while the aerial image is 2D. In addition, applying aerial stereo cannot preserve the required accuracy since its precision is limited by poor aerial triangulation. In this research, we generate a 3D reference for the matching by fusing aerial image with the ALS data. In this method, first the digital elevation model (DEM) is generated from the aerial point cloud. Then, the DEM is applied to the aerial road markings to generate 3D references for the matching. **Figure 4-1** shows the 3D calibration flowchart. 3D registration of MMS data can also remove the roll and pitch offset of the sensor.

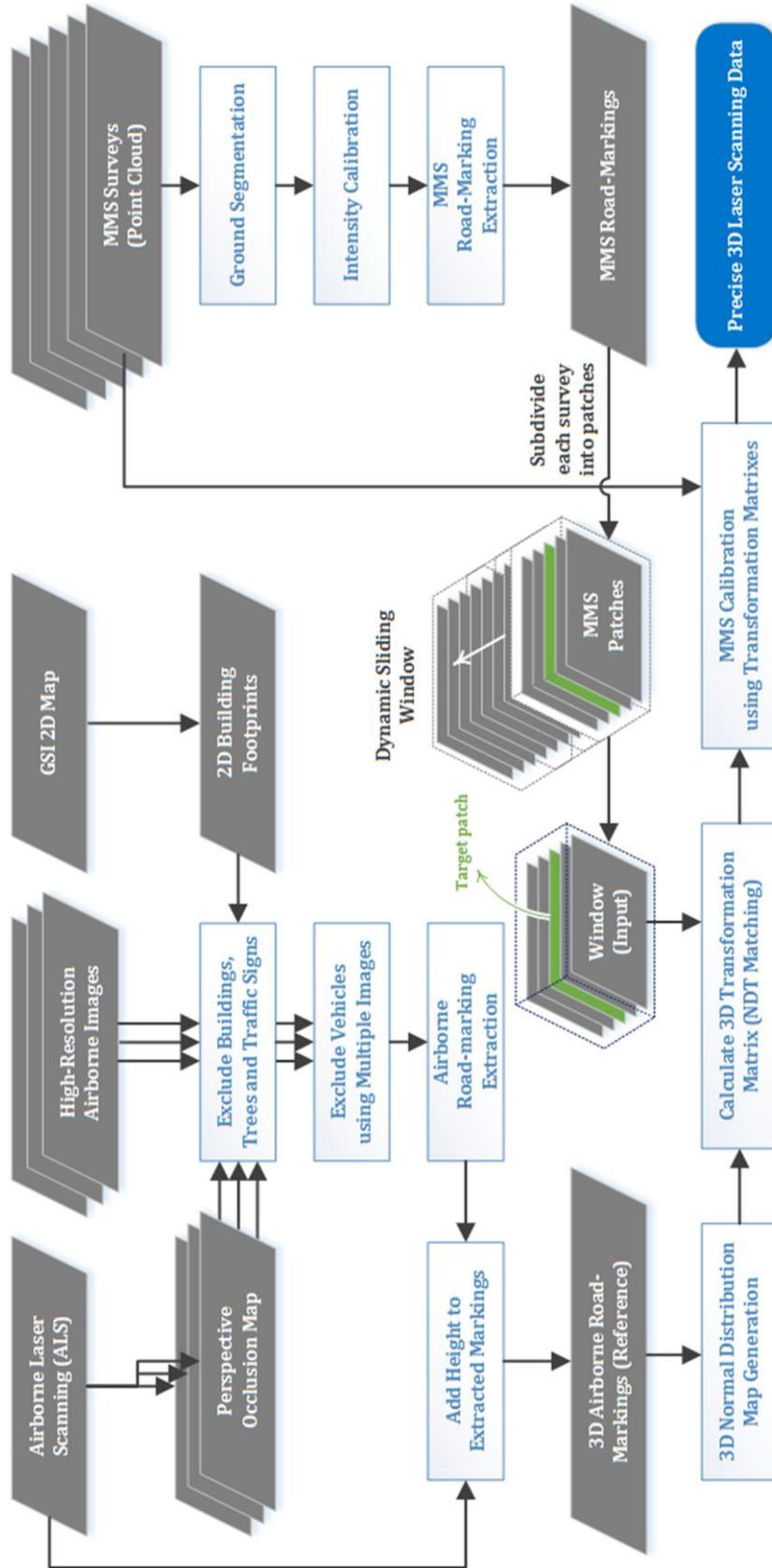


Figure 4-1. The overall workflow of the 3D MMS calibration framework.

## 4.1. Generating 3D NDT map from aerial reference

Extracted road markings from the aerial image have a 2D format. To maintain the 3D accuracy of the MMS, we need to generate a 3D reference by adding precise height information to the aerial road markings. Therefore, we make a high-resolution ground elevation profile from the aerial point cloud. By adding the corresponding height information to the each road marking pixel, we can produce 3D reference road marking.

**Figure 4-2** shows the ground elevation profile generated for the Hitotsubashi intersection.



**Figure 4-2.** Generated high-resolution ground elevation profile from ALS.

## 4.2. 3D MMS registration

In this step, extracted road markings from the MMS survey should be aligned to the 3D aerial reference. For the registration, first a 3D normal distribution map is generated from the aerial reference. Then, MMS road markings of each survey are patch-wise cropped, and points within a dynamic sliding window over the patches are registered to the 3D NDT map. Finally, the original MMS point cloud is calibrated using the calculated transformation matrix.

To register the MMS survey to the aerial reference, we use the point-to-distribution variant of normal distributions transform (P2D-NDT) algorithm [79]. The basic idea of this method is to register the input point cloud over the distribution rather than directly registering to the reference point cloud. First, the target space is regularly subdivided into the fixed-size grids. In our application, we assume the 3D aerial reference as the target point cloud. For each grid, 3D normal distribution of the points is calculated using the following equations.

$$\mu = \frac{1}{n} \sum_{i=1}^n p_i, \quad (1)$$

$$\Sigma = \frac{\sum_{i=1}^n (p_i - \mu)(p_i - \mu)^t}{n-1}, \quad (2)$$

The probability of observing a lane marking at 3D-point  $p$  in this grid is now shown by the normal distribution (3):

$$P(p) \sim \exp\left(-\frac{(p-\mu)^t \Sigma^{-1} (p-\mu)}{2}\right), \quad (3)$$

The grid size is critical for a robust and accurate matching. Usually, larger grid size is more robust and smaller grid size is more accurate. In our method, we use the coarse-to-fine strategy (2.0m to 0.5m grids) for the map-matching.

Each MMS survey is generated from thousands of scan lines. These scans are

registered based on the fusion of GPS, IMU, and odometer to form a complete point cloud. If the self-positing system performs perfectly, points of each laser scan will be registered to their exact global position and generate a precise MMS point cloud. Our goal is to register each scan line to its correct position in the presence of the error. However, points of a single scan do not provide enough features for the registration. Owing to the IMU and odometer, the relative position of consequent scan lines within a small patch (e.g. 0.5m length) is precise. Therefore, we subdivide each survey into small fixed-size patches and consider them as a unit of the registration. Since these small patches still do not have enough road features especially for the longitudinal matching, we propose a dynamic-size sliding window around the target patch overcome this problem.

The problem of the static-size sliding window is that the window length may be too short or too long. If it is short, it may not be able to provide enough feature for a precise matching. On the other hand, the large window may contain a significant relative error in itself which makes the registration difficult. Therefore, we define the length of the window dynamically based on the number of road marking points within the window. Although the registration is done based on the window, the transformation matrix acquired from every matching is applied only to the center patch of the window. Then, the sliding window shifts to the next patch and the window's length may change. In this research, the length of the patches is defined 0.5m based on the IMU performance to limit the deformation inside the patch within a centimeter. The window size is defined dynamically to have at least 30 patches with enough road markings for the registration.

Uneven distribution of the points within the window can cause a considerable error in the NDT matching. Since MMS point cloud is uneven, we have to perform a subsampling before the matching. The best subsampling grid size is obtained empirically

to be 30cm.

For the registration, the input point cloud is moved over the reference NDT map using 3D transformation matrix  $M_{t\_3D}$  to calculate the optimum transformation which obtains the highest matching score.  $M_{t\_3D}$  is defined as:

$$M_{t\_3D} = \begin{bmatrix} R(\phi_x, \phi_y, \phi_z) & T_{x,y,z} \\ 0 & 1 \end{bmatrix}, \quad (4)$$

$$S = \prod_{k=1}^n P(T(M_{t\_3D}, \vec{w}_k)), \quad (5)$$

In the optimization process, calculating the Hessian matrix and gradient from  $S$  is difficult. Therefore, we use the log likelihood (6) as score function.

$$-\log S' = - \sum_{k=1}^n \log(P(T(M_{t\_3D}, \vec{w}_k))), \quad (6)$$

In this case, the best transformation matrix is the one that makes the smallest value for the Eq. (6). However, this value is very sensitive to outliers. Because, outliers make the value of  $P$  very small. Thus, the logarithm of  $P$  become very large, which affects the matching. Therefore, rather than directly using the Eq. (3), the mixture of normal distribution and a uniform distribution is used. This distribution is defined as:

$$\hat{P}_j = \xi_1 P_j(\vec{w}) + \xi_2 P_0, \quad (7)$$

Where  $P_0$  is the expected rate of the outliers, and  $\xi_1$  and  $\xi_2$  are constants such that integral of  $\hat{P}_j$  become 1. Eq. (5) is approximated by a Gaussian function and we consider that function as score. For more details about the map-matching using optimization problem, please refer to [79].

### 4.3. Experiments and evaluations

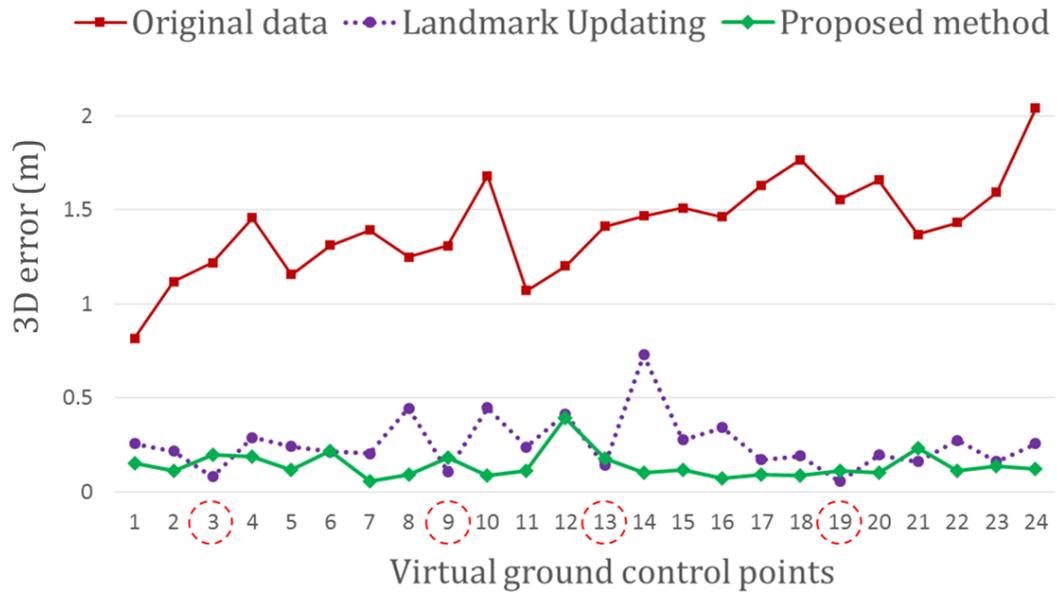
To evaluate the performance of the proposed method, we measured seven different 3D GCPs at the corner of traffic signs using RTK-GPS and total station. Then, 24 virtual control points (VCP) are generated by propagating the original GCPs. Using

these VCPs, the performance of the conventional MMS, landmark updating and proposed were assessed. The accuracy of each method is shown in **Table 4-1**. We can see that the proposed method could decrease the average 3D error of the MMS data from 1.47m to 0.13m. The proposed method could also outperform expensive and labor-intensive landmark updating method. The average error of the landmark updating was 26cm whereas the maximum error was 71cm. **Figure 4-3** shows the error for all 24 VCPs and **Figure 4-4** shows the evaluation of 3D registration visually.

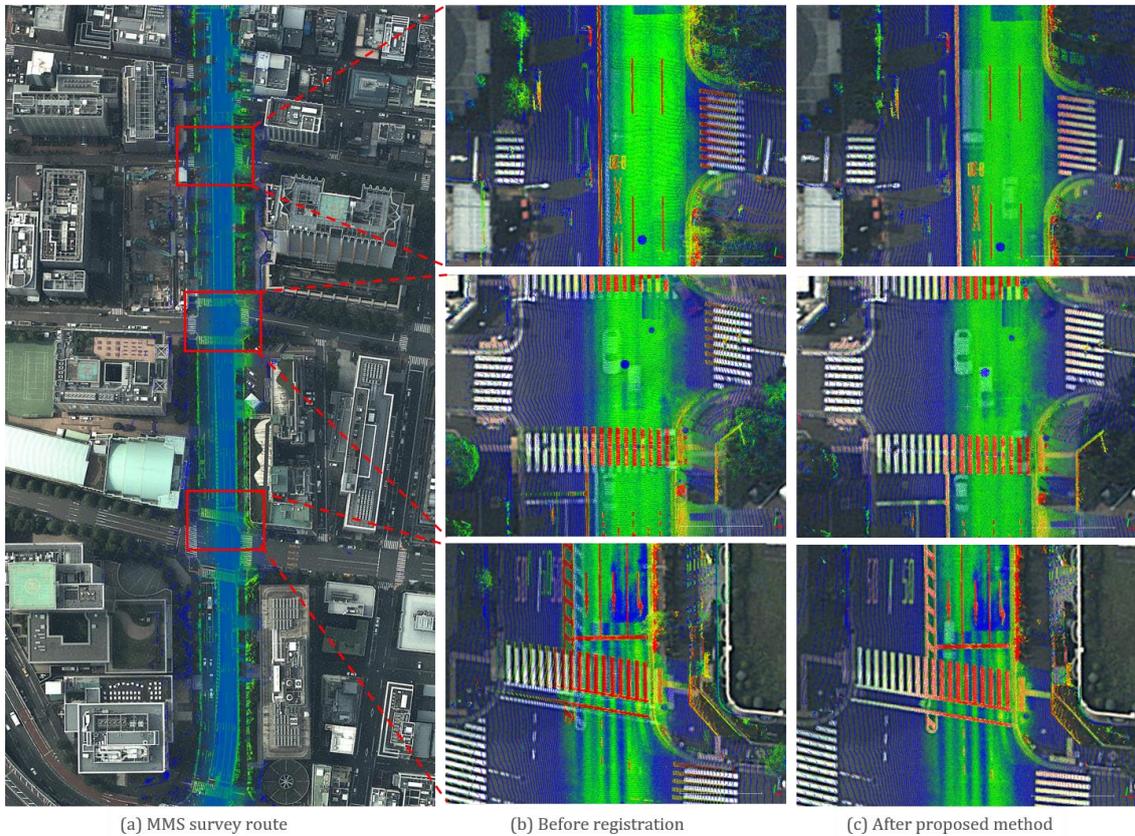
As we can see in **Figure 4-3**, the landmark updating method over performs our proposed method only in the points close to the GCPs. In this figure, virtual control points 3, 9, 13 and 19 are relatively close to the GCPs used for the landmark updating. Since the landmark updating method adjusts the MMS data only based on a few GCPs, it has a good performance for the area around GCPs. However, the error model in other points which have few meters distance to GCPs follows the error model of the original data.

**Table 4-1.** Total error of the nine surveys

Method	Error (m)		
	Mean	Max	Std
Original MMS data	1.47	2.07	0.26
Landmark updating	0.26	0.71	0.10
Proposed method	0.13	0.39	0.07



**Figure 4-3.** Error of the MMS laser scanning data before calibration, after the 3D calibration using the landmark updating method, and after applying the proposed methods. The marked virtual control points are relatively close to the GCPs measured for the landmark updating method.



**Figure 4-4.** Visual evaluation of the proposed 3D registration method: (a) Survey route of the MMS on the aerial image; (b) Enlarged view of the original MMS point cloud on the aerial image; (c) Enlarged view of the MMS point cloud after applying the proposed method.

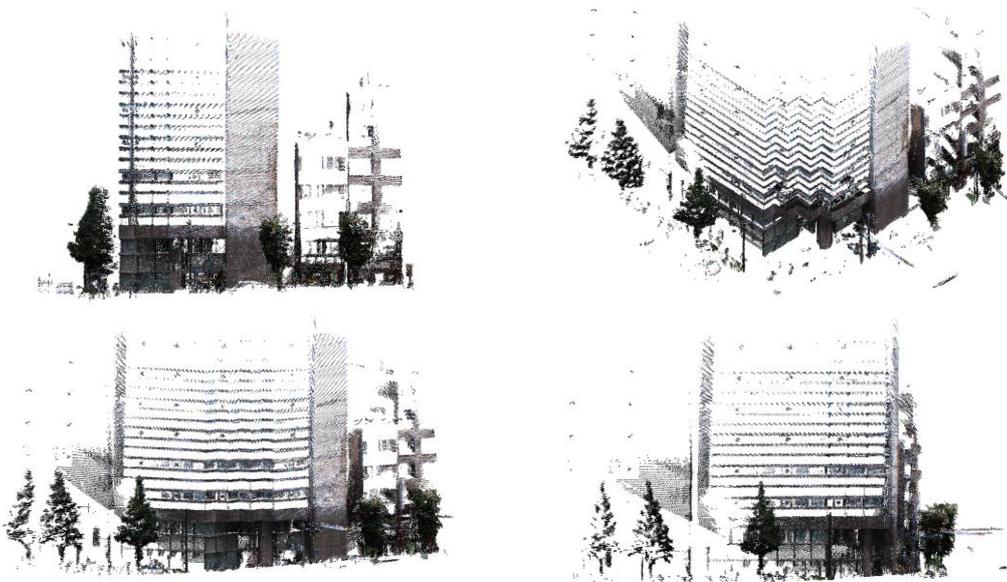
For the VCPs which have few meters distance with GCPs, our proposed method has much higher performance. In overall, the proposed using the proposed method, we can achieve a high-precision 3D urban mapping without requiring labor-intensive GCP measurement.

**Figure 4-5** shows a building after 3D registration of MMS scans. It can be seen that the misalignment before the calibration almost removed.

## 4.4. Conclusions

In this chapter, we proposed a complete framework for the automatic 3D registration of the MMS laser scanning data to the aerial surveillance data, which could improve the accuracy of the mobile mapping system significantly. For this purpose, 3D road markings from the combination of high-resolution aerial image and Aerial Laser Scanning (ALS) were considered as a reference for the registration.

The experimental results of the Hitotsubashi intersection in Tokyo has been used to evaluate the performance of the proposed method. The results demonstrated that the



**Figure 4-5.** Point cloud of one building after 3D registration. As can be seen misalignment removed.

proposed method is practical for the precision 3D mapping of the urban area and it could achieve a mean error of 13cm in a challenging scenario.

## **Chapter 5.**

# **3D Urban Modeling Using Ground and Aerial Surveillance Data**

### **5.1. Introduction**

Research on autonomous vehicles and self-driving cars have become one of the popular topics in Intelligent Transportation Systems (ITS) and many research projects have been launched toward achieving the goal of driving without human supervision over the last few years. The attention of the research in autonomous driving has recently switched its focus from the well-structured environments in highways to more unstructured environments, like urban traffic scenarios [87], [88].

One of the key requirements of such a vehicle which drives without human input is precise self-localization, especially in dense urban areas. Accurate and reliable self-positioning will let the vehicle to be guided along the desired trajectory toward the pre-defined destination.

Global Navigation Satellite System (GNSS), stereo-vision based visual odometry and scan matching using 360 degrees light detection and ranging (LiDAR) scanners are three main self-localization techniques for autonomous vehicles.

GNSS is a proven vehicle self-localization technique in the open sky. However, it suffers from the effect of multipath and Non-Line-Of-Sight (NLOS) propagation in dense urban areas [89]. There are different techniques which use the 3D map to detect and mitigate the multipath effects [90]–[92]. These methods called 3D map aided GNSS

positioning have recently gained more attention because they do not require any extra hardware besides GNSS receiver. In [93], [94], 3D building model and the ray-tracing algorithm is used to rectify multipath and NLOS effects. In such a 3D map aided GNSS positioning technique, accuracy and the level of details of the 3D building model play a major role in detection and rectification of Multipath and NLOS.

Another vehicle self-localization technique uses stereo-vision based visual odometry. Visual-odometry, which estimates the full six-degree of freedom (DOFs) of vehicle motion from image sequences, produces relatively accurate and has lower drift than inertial measurement units (IMU). Visual-odometry methods are, however, incremental and prone to small drifts, which, when integrated over time, become increasingly significant over large distances [95]. One of the solutions for minimizing drift of VO is to use a precomputed 3D model of the environment. Localization with respect to the 3D model remains drift-free [96].

360-degree LiDAR can generate a dense point cloud and self-localization using scan matching acquires centimeter-level accuracy. However, scan matching using a precise and priori known high-resolution point cloud map [88], [97] is not feasible for urban traffic scenarios. Assume the scenario that many autonomous cars are running simultaneously. If each vehicle needs to download a large high-resolution 3D point cloud map for localization, network congestion will be a big problem. Therefore, a compact 3D building map which provides geometry of surrounding buildings can be replaced with large 3D point cloud map.

As described above, an accurate city scale 3D map can assist major vehicle localization techniques.

Besides assisting self-localization in ITS, 3D building map plays a key role in

wireless base station placement [98], wind simulation, air pollution simulations, noise propagation estimation [99], urban planning [100], etc.

With the development of the laser scanning technology, airborne laser scanning (ALS) has now become readily available and cost effective for the acquisition of high-resolution digital elevation models. However, most of the methods require human interaction with the software and still modeling results are not simplified enough while preserving important structural features [101]. In this chapter, an automatic 3D building map reconstruction technique is proposed which integrates ALS point cloud with 2D boundary map. 2D boundary map is used to assist the building segmentation of the point cloud in dense urban areas, in which defining building edges between two connected buildings are difficult.

## **5.2. Related works**

Since a broad range of applications benefits from 3D building map of cities, the research on 3D building map reconstruction has considerably progressed in the last few years. Various technologies ranging from computer vision and image processing, photogrammetry and remote sensing have been proposed for this purpose.

Many researchers with the background of computer vision and image processing developed new techniques and algorithms to detect roof edges and reconstruct city scale 3D building model. However, roof edge detection using aerial images are not efficient enough for dense urban areas in which buildings are connected to each other and the edges of the each roof is not clear.

Many other types of research are using point cloud from remote sensing techniques such as laser scanning for building reconstruction. The first step of building reconstruction from LiDAR point cloud is the extraction of individual buildings' points

from a point cloud. Once buildings are extracted, there are two different approaches for 3D building model reconstruction: model-driven approaches and data-driven approaches.

Model-driven approaches use a group of pre-defined models and choose one of the models for each building or part of each building [102]–[104]. Model-driven reconstruction of buildings is fast and efficient if the model of building exists between the predefined models and has been selected correctly, because regularization constraints such as parallelity and orthogonality of the edges are already inherent in the model. However, model-driven approaches are limited to the few number of pre-defined models and therefore they are not flexible and applicable for dense urban areas in which buildings have variety in shape and complexity [105].

In contrast, Data-driven approaches connect individual roof segments, which are constructed according to a preliminary segmentation of the building point cloud. Compared to Model-driven approaches, Data-driven approaches require a high effort for regularization and improving the results. Even though they are widely used [106], [107]. The advantages of these approaches are high fit to the input data and flexibility in modeling unknown building models or complex roof shapes. One of the challenges in model-driven approaches is roof extraction and roof reconstruction. Since these methods require extraction of roof features directly from point cloud data, they need more computation than model-driven methods. These methods achieve roof segmentation by using different techniques such as line and surface fitting techniques such as RANSAC, Hough Transform, or using region growing methods.

### **5.3. Proposed method**

**Figure 5-1** illustrates the workflow of the proposed 3D building map

reconstruction framework. The input data consist of airborne laser (LiDAR) scan 3D point cloud data and 2D building boundary map from The Geospatial Information Authority of Japan (GSI). 2D map provided by GSI or most of the national organizations only contains a rough footprint of all buildings in the city. Therefore, the 2D map can ease building segmentation by providing an acceptable estimation of building boundaries. The proposed framework consists of 4 major steps for each building: (1) the first step is building segmentation assisted by 2D boundary map; (2) the second step segments out roof segments from top to bottom; (3) the third is roof shape reconstruction which contains three sub-steps; and (4) the last step reconstructs building model using information from step 2 and 3; In the following each step of the proposed method will be described in detail.

### **5.3.1. Building segmentation**

The first step of the proposed framework is building segmentation. In building segmentation, the set of all points representing a building should be extracted from ALS point cloud. This step has an important effect on the computation cost of methods with dense point clouds. The proposed method hires 2D map provided by The Geospatial Information Authority of Japan (GSI) for building segmentation. 2D map contains rough footprint of all buildings. Although they are not accurate, provided footprints are giving a good estimation of the building boundary in 2D. Since the initial 2D map is not accurate, if it is used directly for building segmentation, some important boundary points of the building can be mistakenly excluded from set of points which are segmented out as a building. **Figure 5-2** shows this problem. Therefore, building segmentation in proposed method is performed by first expanding building map by the pre-defined threshold, then extracting all points within the expanded boundary. The next step is roof segmentation.

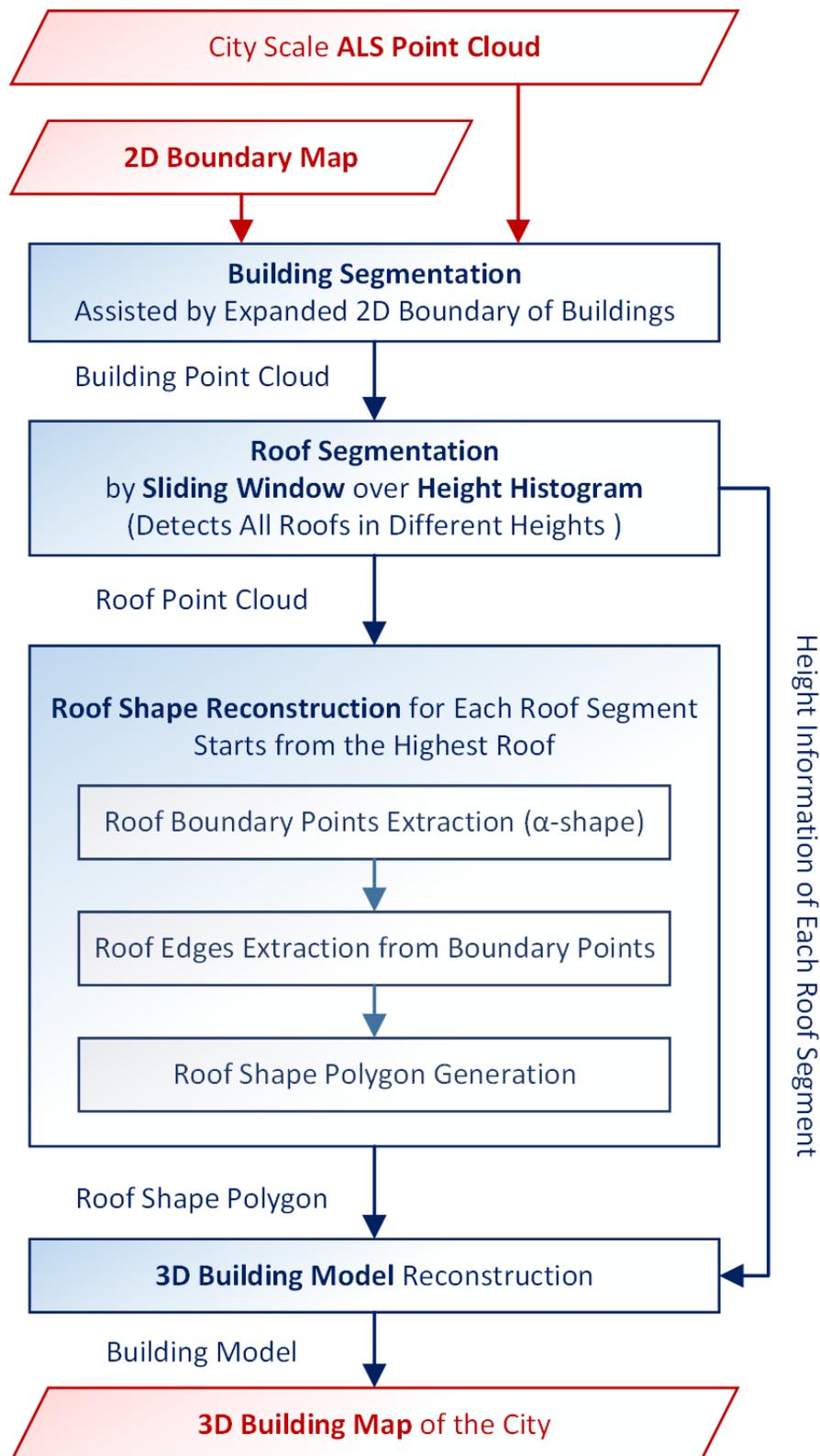
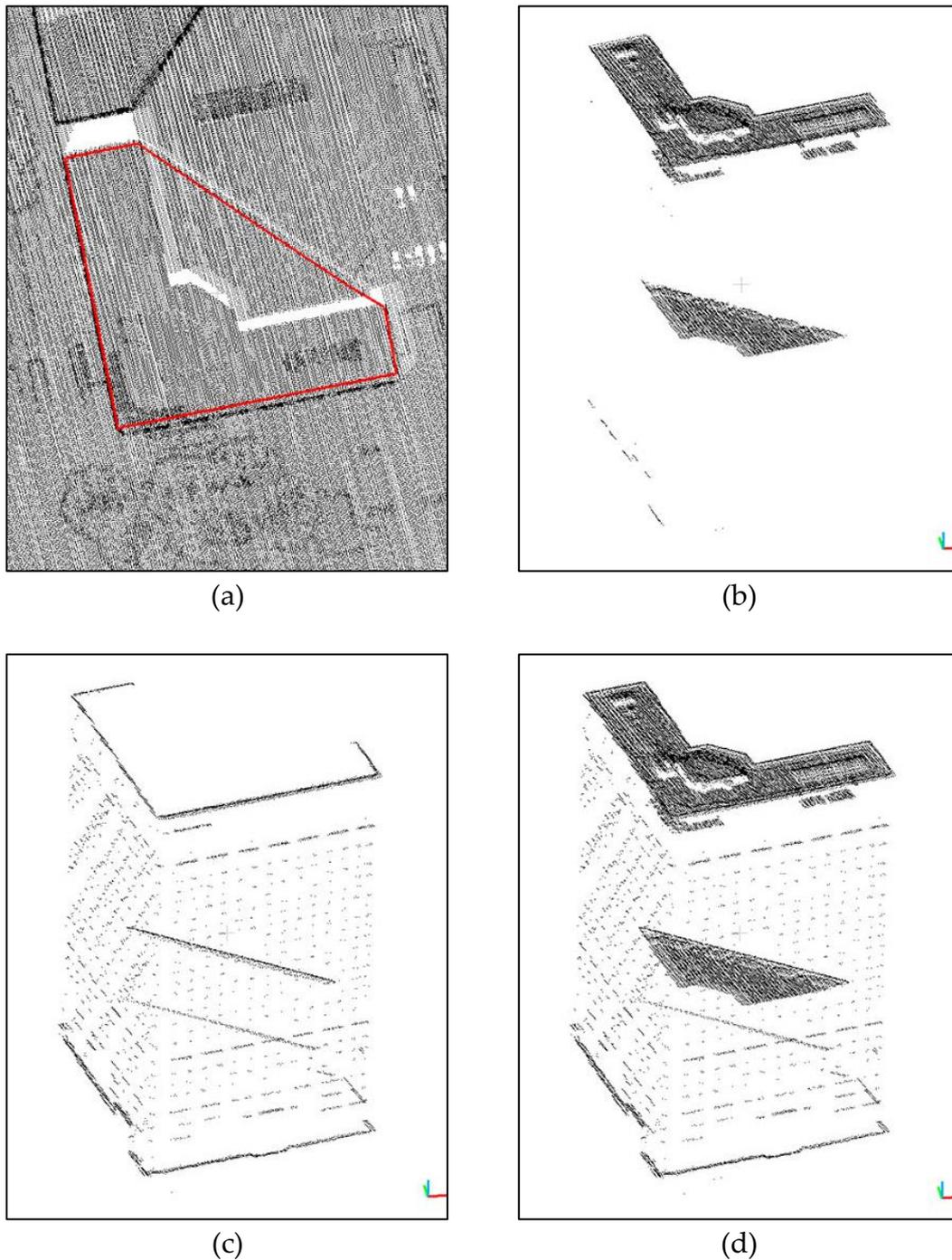


Figure 5-1. Workflow of the proposed framework



**Figure 5-2.** Building segmentation by initial 2D map. (a) Top view of the ALS point cloud. GSI 2D map is shown by red polygon. As shown, 2D map only provides a footprint of the building and it cannot represent the shape of building considering multiple roofs. (b) Extracted set of the building segment points using initial 2D map. Important boundary points of the building are missing because 2D map is not accurate enough. (c) Boundary points of the building which are mistakenly excluded from building segment. (d) Set of building points extracted using expanded 2D map with pre-defined threshold.

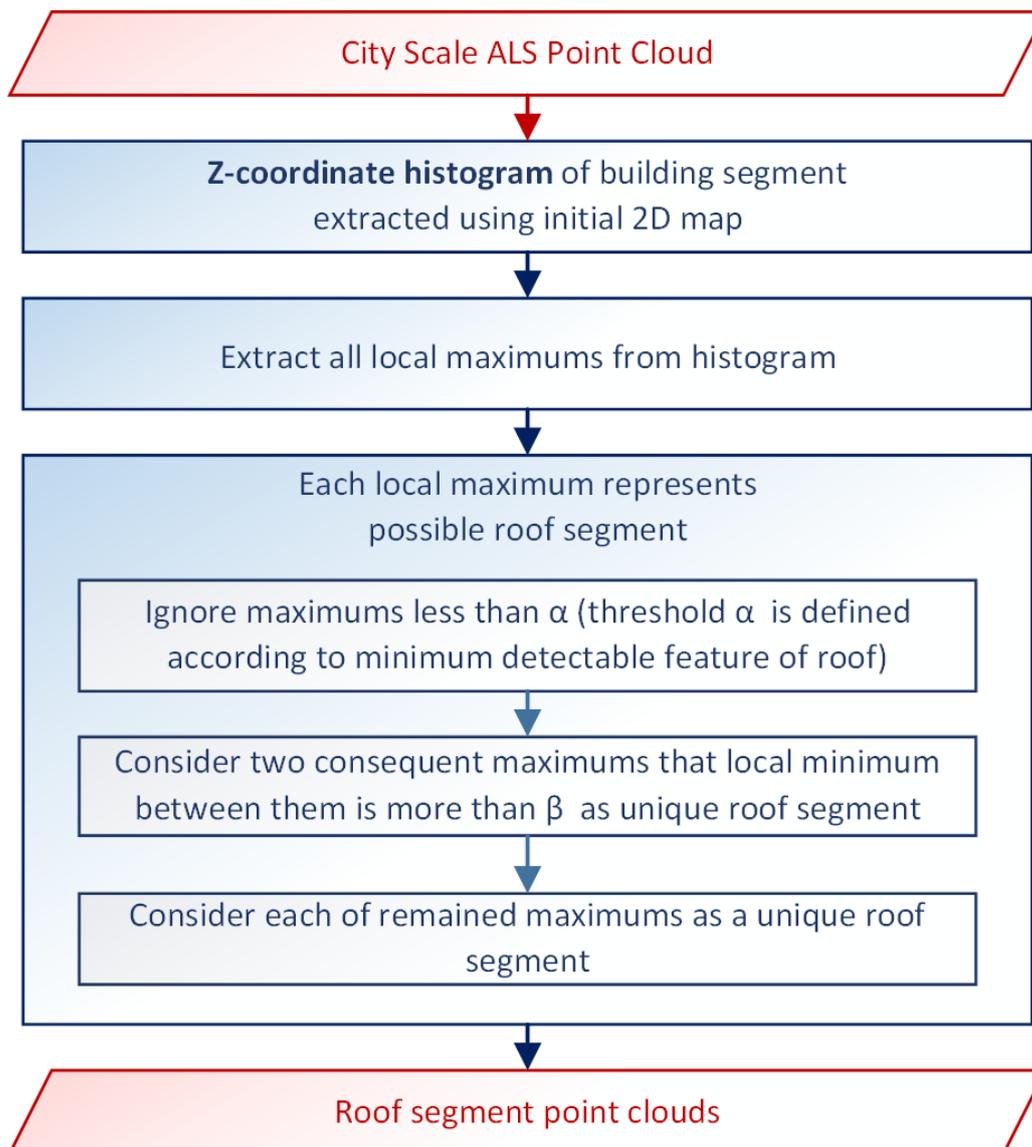
### 5.3.2. Roof segmentation

2D map provides only a footprint. Therefore, a building segment can consist of multiple roofs. In this step, all points belonging to each single roof of the building should be extracted. For segmentation of multiple roofs in a building, a method using height histogram of a set of building points that is extracted using primary 2D map is proposed. In this method, the local maximums of the histogram represent the height of possible flat roof segment.

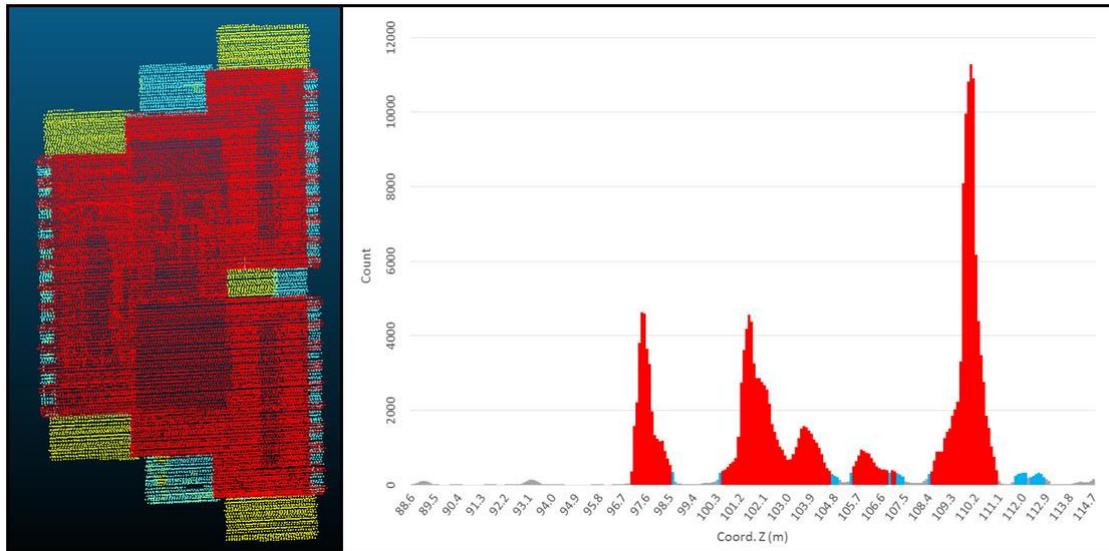
For roof segmentation, instead of an expanded 2D map, the primary 2D map is used. Building segment extracted using expanded 2D map usually contain ground or non-building points. However, building segment extracted using primary 2D map usually contains only building points. Therefore, using primary 2D map makes roof segmentation more accurate. In this approach, only flat roofs are reconstructed. Using this method, non-horizontal (pitched) roofs also can be modeled using flat planes.

For roof segment generation, first height histogram of building point cloud is calculated. For detecting the height of each roof segment accurately, overlapping bins are used. This means the width of the histogram bin is set as 50 cm but two consequent bin has only 10 cm shift. Therefore consequent bins are 80% overlapped. After histogram is calculated, all local maximums of the histogram are considered as possible roof segments. Then, all local maximums that their value is smaller than threshold  $\alpha$  are ignored.  $\alpha$  is the minimum number of points that represent one roof segment. The minimum size of the detectable roof is the main factor for defining  $\alpha$ . By multiplying the size of the minimum detectable roof  $s$  ( $m^2$ ) with  $d$ , the density of the scan per  $m^2$ , the minimum number of points  $\alpha$  can be calculated. In our method, the minimum size of the detectable roof is defined as  $50m^2$ ,  $d$  is around  $7 \text{ pts}/m^2$  and therefore  $\alpha$  is calculated as 350 points.

After defining all local maximums more than  $\alpha$ , all consequent local maximums that the local minimums between them are more than threshold  $\beta$  are marked as unique roof segment. Usually, flat roofs have a border belt called parapet which it is slightly higher than the roof itself. This structure usually makes such consequent local maximums. In our method, roof and parapet are reconstructed as a single roof. In this case, the height



**Figure 5-3.** Workflow of the roof segmentation



**Figure 5-4.** Roof segment detection using z-coordinate histogram. Red circles are possible roof segments.

of roof will be equal to the height of the parapet. In the final step, each of the remaining local maximums will be considered as a unique roof segment. For each roof, maximum z value between all its points is assigned as the height of the roof.

### 5.3.3. Roof shape reconstruction

The third step is roof shape reconstruction from a set of roof points. Roof shape reconstruction consists of 4 sub-steps: 1) filtering outlier points; 2) extraction of boundary points; 3) extraction of roof edges; 4) reconstruction of roof polygon.

- 1) In order to keep only the points that are representing the roof surface, outlier points should be filtered. For this purpose, radius outlier removal filter is used which removes all points from input cloud that don't have at least pre-defined number of neighbors within a certain range. The number of neighbors depends on the density of input point cloud.
- 2) Next step is extracting the set of points that represents the outer boundary of each

roof. The shape of a roof is not necessarily convex. Therefore, we use the  $\alpha$ -shape algorithm for defining this set of points. Given a set of points, the  $\alpha$ -shape is essentially defined as the subset of the Delaunay triangulation of the points where the Delaunay cells' radius is below positive real number  $\alpha$ .

- 3) In the third step, roof edges should be defined using boundary points.  $\alpha$ -shape of each roof segment consists of less than few hundreds of points. Therefore, Hough Transform is suitable for feature extraction and edge detection. Consider that in the  $\alpha$ -shape, points are sorted in order. Defining all possible roof edges consists of the following steps:

1. Define the best fitting line using Hough Transform.
2. Stop when the number of points on the detected line is less than  $n$ .
3. Remove all consequent points that their distance to the line is less than  $d$  from  $\alpha$ -shape excluding the first and last point of each sequence.
4. Store the first and last points from the Step 3 for the detected line.
5. Go back to Step 1.

As mentioned before, the number of points in  $\alpha$ -shape of a roof is limited, and sometimes an edge is represented by only 5 or 6 points. Also, a corner point can represent two different edges of a roof. In this case, if in previous iterations of the algorithm a corner point was removed because of being close to one the lines, then the quality of the other line of that corner will be decreased in Hough Transform. In order to overcome this problem in Step 3, first and last points of each sequence are excluded since they can be a corner point with a high possibility. The reason which only consequent points are removed is that non-

consequent points are possibly not on edge but it is close to a detected line just by chance.

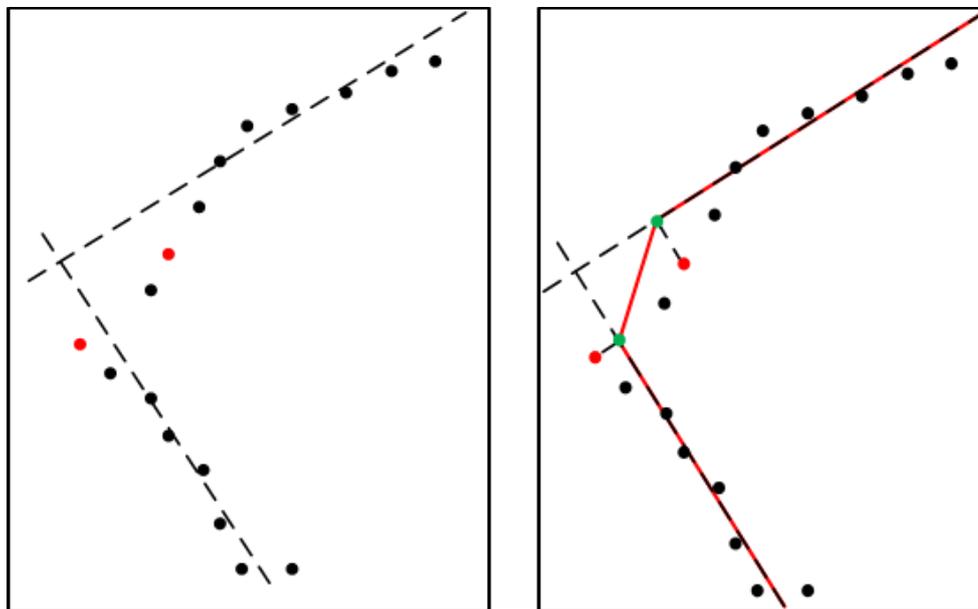
- 4) In the previous step, all detected lines in addition to first and last points of a corresponding line are stored in an array. In the last step, a closed polygon that represents the roof boundary should be reconstructed using lines and their first and last points. Roof shape reconstruction is consisting following steps:

1. Start from the first detected line  $l_0$ .
2. Add the first point of  $l_0$  to `roof_polygon`.
3. If the last point of `current_line` is same with the first point of another line, select that line as `next_line`. If not go to Step 5.
4. Add the last point of `current_line` to `roof_polygon` and Go to Step 8.
5. Search a line with a closest first point to the last point of `current_line` and set it as `next_line`.
6. Add the perpendicular image of the last point of `current_line` to `roof_polygon`.
7. Add perpendicular image of the first point of `next_line` to `roof_polygon`.
8. Set `current_line` as `next_line`.
9. Go back to Step 3.
10. Stop when no line remains.

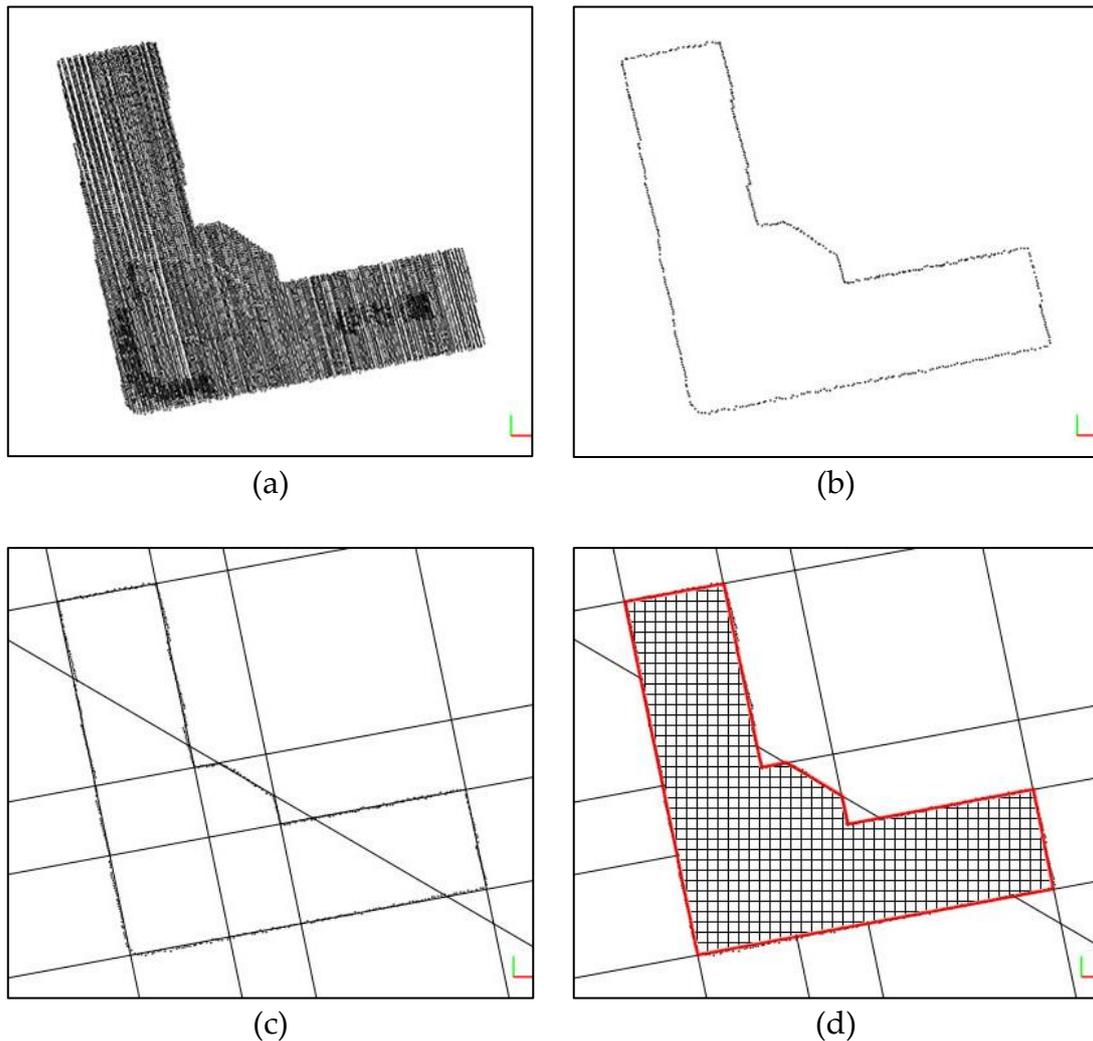
There are two cases in roof polygon reconstruction. In the first case, the last point of a detected edge is same with the first point of another detected edge. In this case, roof polygon is simply reconstructed by connecting these two lines (Steps 3 and 4). In the second case, the last point of a detected edge is not matching with the first point of its consequent edge. **Figure 5-5**(left) shows this phenomenon. In this case, roof polygon is reconstructed by connecting the perpendicular image of those points to each corresponding line (Steps 5-7) as shown in **Figure 5-5**(right). **Figure 5-6** illustrates the flow of roof shape reconstruction step by step.

### 5.3.4. 3D building model reconstruction

After completing roof shape reconstruction for all roof segments, 3D model of the target building can be reconstructed by connecting roof polygons in their corresponding heights together. By repeating previous steps for all buildings in 2D map



**Figure 5-5.** Roof polygon reconstruction when neither first nor last point of two consequent lines are matching. Black points are points in the  $\alpha$ -shape of the roof. Dashed lines are detected lines by Hough Transform. Red points are first and last points of each line. Green points are perpendicular image of the first or last point on the corresponding line. Red line is a part of reconstructed roof polygon.



**Figure 5-6.** Roof polygon reconstruction result. Black points are points in the  $\alpha$ -shape of the roof. Dashed lines are detected lines by Hough Transform. Red points are first and last points of each line. Green points are perpendicular image of the first or last point on the corresponding line. Red line is a part of reconstructed roof polygon.

of GSI, 3D building map of the city can be reconstructed in a short time.

## 5.4. Experimental results

For evaluation of the proposed method, Shinjuku area in Tokyo, where is one of the urban canyons of Japan is considered.

### 5.4.1. Specification of the inputs

Proposed framework receives two input: ALS point cloud data and 2D boundary

map. For ALS, point cloud data provided by ATEC Company is used. This is data is acquired by an airplane equipped with ALS70 LiDAR scanner and GNSS/IMU system. The density of point cloud is around 7 pts/m<sup>2</sup>. For the second input, 2D map, boundary map provided by The Geospatial Information Authority of Japan (GSI) is considered.

#### 5.4.2. Parameters

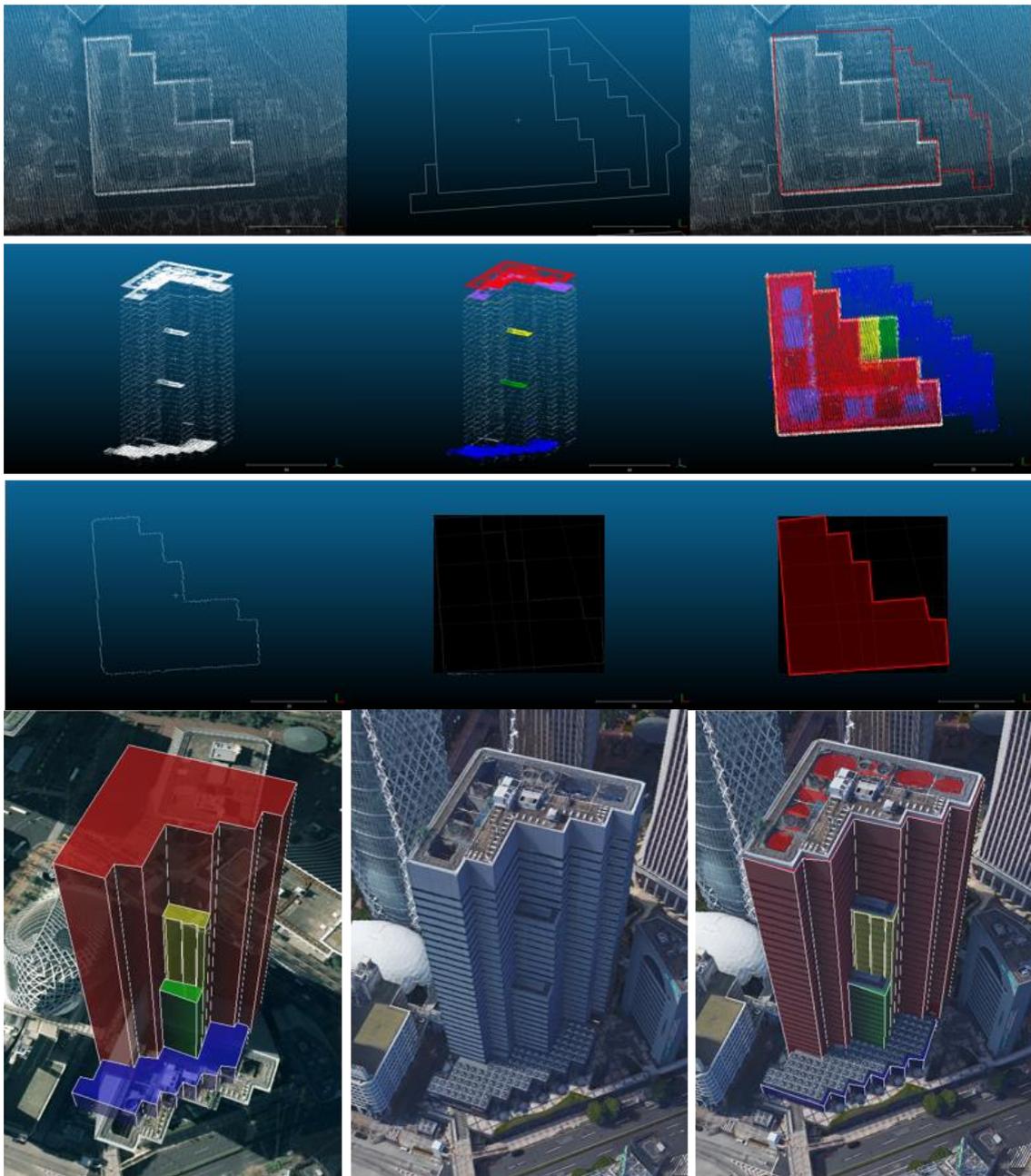
The minimum size of each detectable roof is defined 50 m<sup>2</sup>. By multiplying the size of minimum detectable roof  $s$  (m<sup>2</sup>) with the  $d$ , density of the scan per m<sup>2</sup> minimum number of points  $\alpha$  will be equal to 350 points.

### 5.5. Challenges

Besides benefits there are also pitfalls for 3D map reconstruction using ALS. Since in cities such as Tokyo, limitations of lowest safe altitude for flight are strict, the density of ALS point cloud is less than 10 pts/m<sup>2</sup>. In this case reconstruction of small roofs and edges in complex buildings are quite challenging. On the other hand, ALS point cloud of an area is usually acquired by a single straight scan. In such a scans, there is a high possibility that some edges of the building are perpendicular to the flying direction. For scans that flight direction of the airplane is almost perpendicular to some edges of a building, only limited points can represent corresponding edge, and therefore reconstructed edge will not be a straight line. **Figure 5-8** shows such a situation.

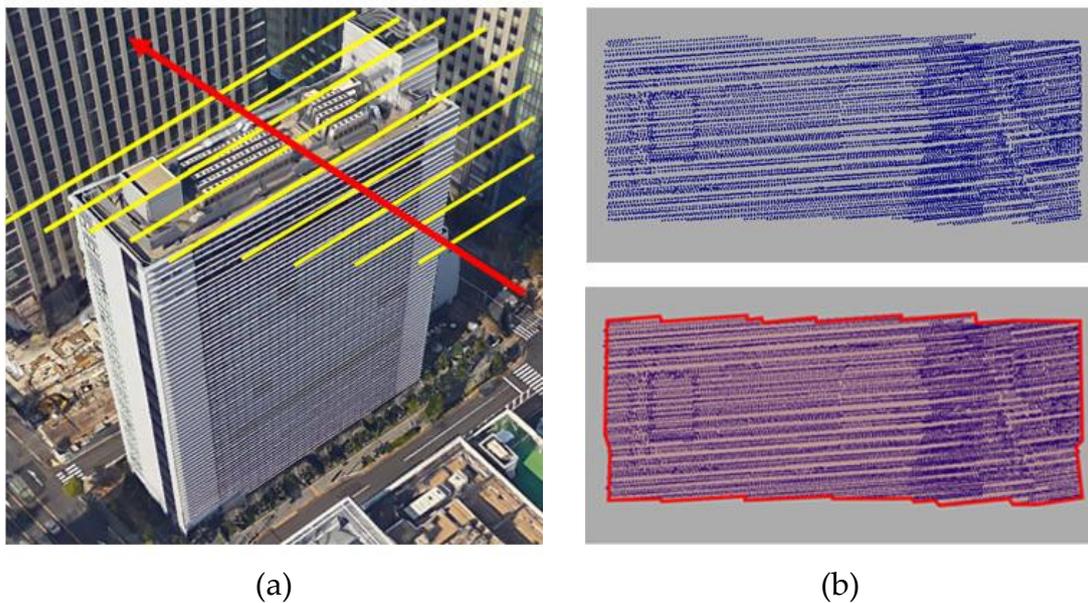
In order to overcome these challenges, different methods can be suggested. The first suggestion is using scan data of Mobile Mapping Systems (MMS) which acquires dense point cloud. Another suggestion is to scan an area multiple times, each time perpendicular to the previous scans, like a grid.

Another drawback of the proposed framework is that it cannot reconstruct



**Figure 5-7.** 3D reconstruction steps for Shinjuku L Tower in Shinjuku, Tokyo.

buildings with curved shapes. Polygonal mesh-based 3D building modeling is able to cover this problem. However, the map size would again increase.



**Figure 5-8.** ALS point cloud for scans that its direction is perpendicular to some building edge; (a) Scanned building and flying direction; (b) Top view of ALS point cloud and reconstructed roof shape in red.

## 5.6. Building wall adjustment utilizing GPS raytracing

As mentioned in the introduction accuracy of building wall can cause an undesired situation in various application. For example, for the Autonomous vehicle self-localization application, the error of the position of the wall can directly cause an error on the positioning. One of the applications of this map is autonomous vehicle self-localization. Other Applications can be affected by the accuracy of the building position as well.

One of the challenges for making accurate wall position of the wall is parapet. In this case, the wall generated by the proposed method is based on the parapet and has error around 1 meter. So after making 3D building model we need to adjust the wall more accurately. To do this, we used the method proposed in [108] this method uses the GPS signal and reflection pass of the signal to rectify the position of the wall. This is shown in

Figure 5-9.

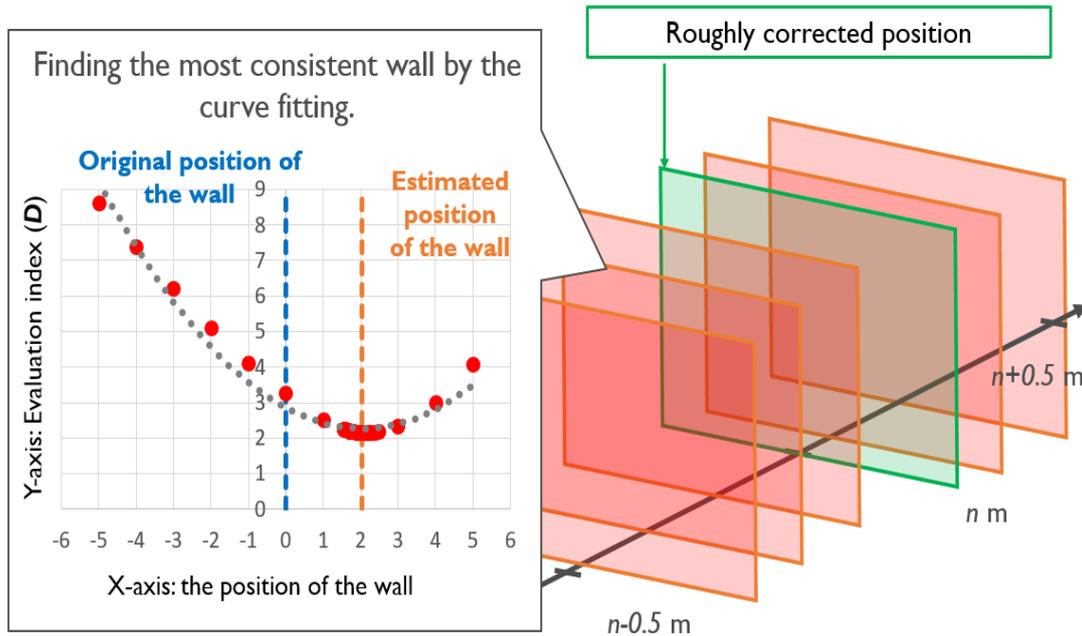
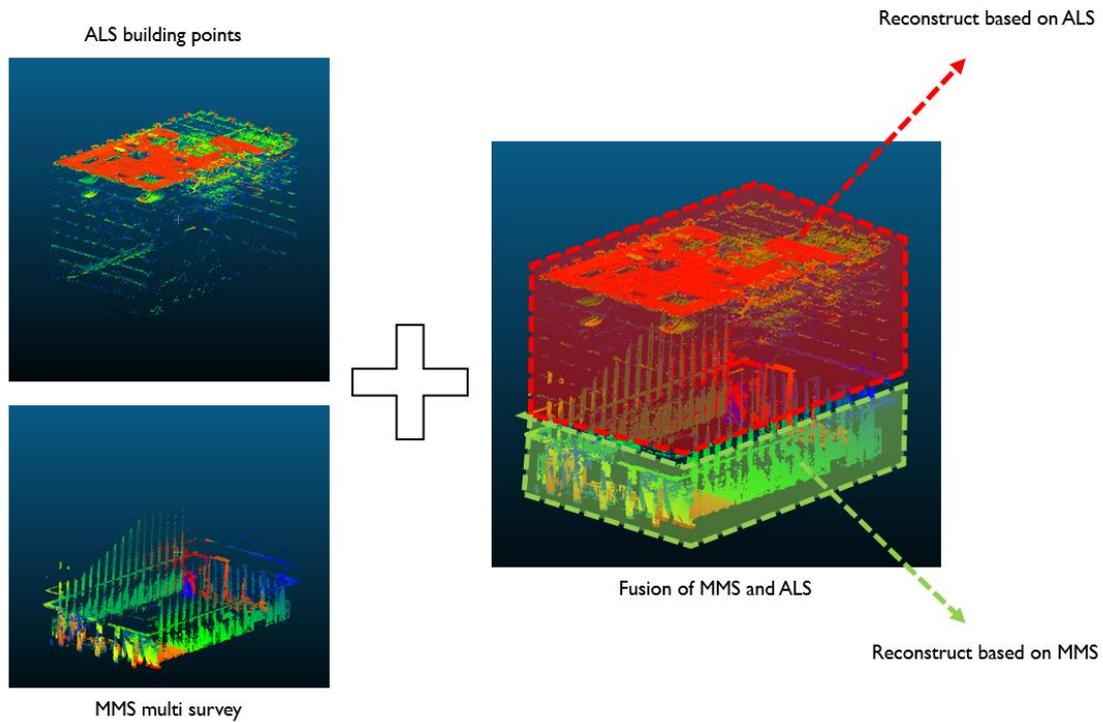


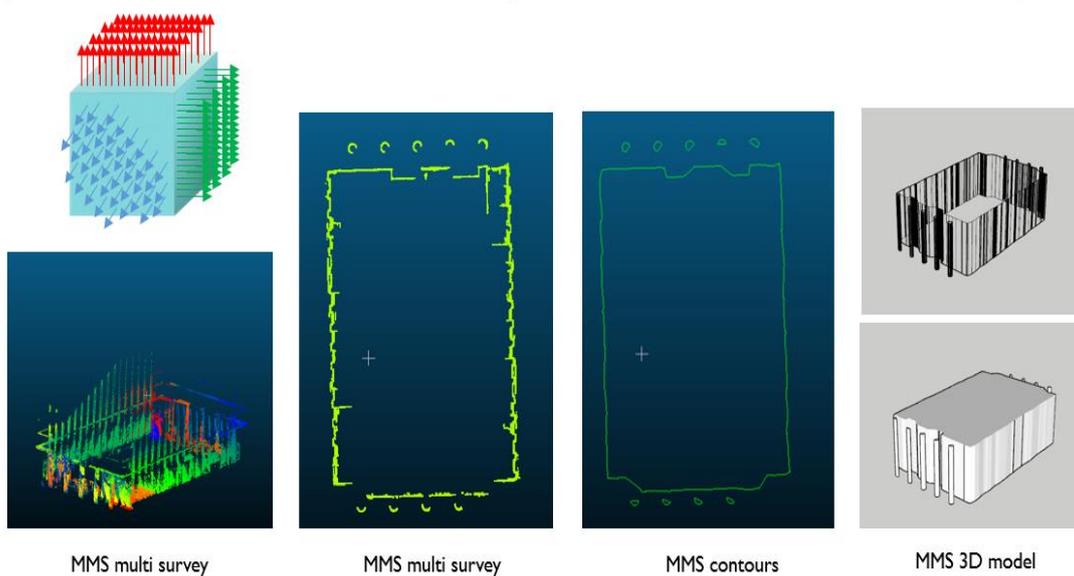
Figure 5-9. Building adjustment using GPS signal and ray-tracing technique

## 5.7. 3D building reconstruction using MMS and ALS

In this section, we focus on making 3D building model from both ALS and MMS. MMS data is good for the reconstruction of the bottom part of the building, and ALS data is suitable for the upper part. **Figure 5-10** shows both MMS and ALS data of a building. As shown in this figure, the ALS covers the roof and upper part of the building and the MMS data covers the lower building walls. For reconstructing building walls, first, we segment out the MMS walls from roofs based on the point cloud normal. Then, we Filter-out outliers using statistical outlier removal. Later, a general segmentation method is used to segment the remaining points. Finally, the concave hull algorithm is applied to the remaining points and the contours are generated for each segment. This work flow is shown in **Figure 5-11**.



**Figure 5-11.** Make 3D building model utilizing ALS data and MMS data sources. In the right figure, the green part shows the MMS data and the red part shows the ALS data.

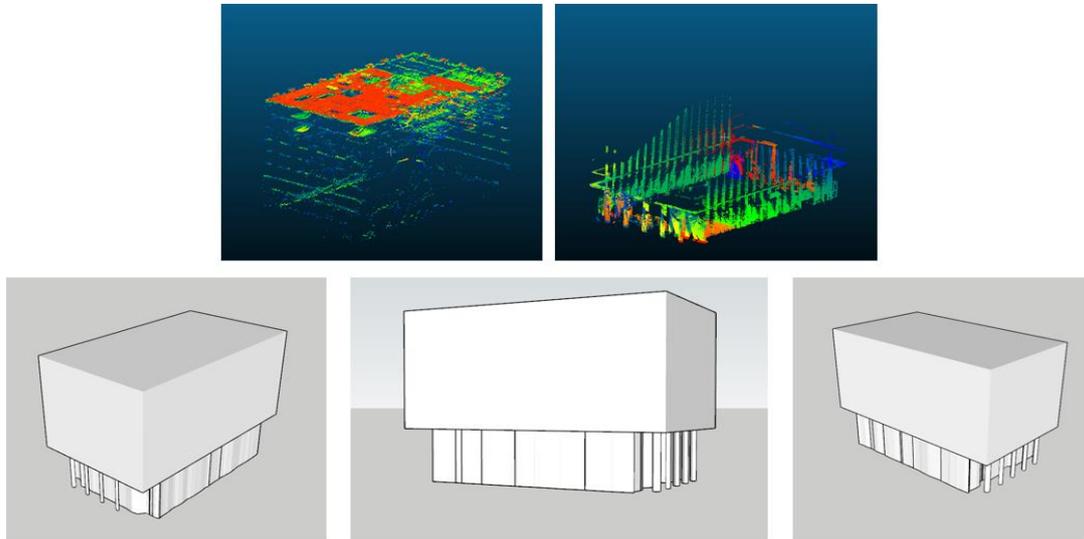


**Figure 5-10.** Workflow of the MMS 3D model reconstruction.

We use MMS data only for the point between ground and a specified height (20m), and the 3D building model of the points higher than the specified height is

reconstructed utilizing the ALS data. Then we easily combine these two data together.

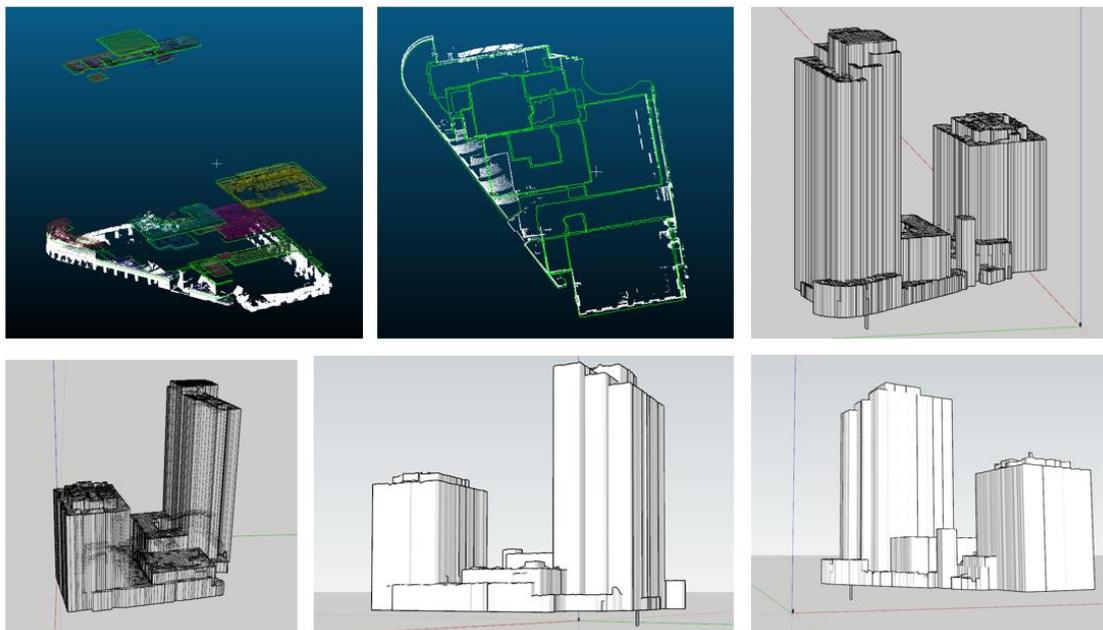
The result of combination of these two building are shown in **Figure 5-12**.



**Figure 5-12.** 3D map reconstruction by fusing MMS and ALS point clouds

It is difficult to reconstruct some complex buildings using only the straight lines.

In this case we can use more flexible structures like 2.5 D meshes. The 2.5 D meshes for



**Figure 5-13.** 2.5D mesh generated from the ALS point cloud for a building in Hitotsubashi area

the one building in Hitotsubashi area is shown in **Figure 5-13**.

# Chapter 6.

## Conclusion and Future Work

### 6.1. Conclusion

In this dissertation, we proposed a thorough framework for the urban mapping which fused aerial surveillance data and mobile mapping data to perform high-definition, high-accuracy and high-precision 3D mapping and modeling of the urban areas. In the proposed framework the aerial data was taken into account for acquiring the global accuracy and the 3D geometry of the structures such as buildings, and MMS was considered for the high-definition data acquisition of the urban area from street view. Aerial and ground-based surveillance platforms have complementary features which were considered effectively in the proposed framework. Finally, the GNSS measurements were employed to analyze and optimize the 3D position of the walls in the generated building models.

In the second part of the dissertation, we proposed a novel framework to generate high-definition mobile mapping data which met the requirements. The proposed system contained these steps: 1) Acquiring aerial data consisting of high-resolution aerial image and aerial laser scanning (ALS) with the fusion of high-end RTK-GPS and IMU; 2) Generating high-resolution mobile mapping surveys containing camera and LiDAR point cloud using high-end RTK GPS, IMU and odometer; 3) Calibration of the aerial data with few number of ground control points (GCPs); 4) Filtering the buildings and their relief

displacement in the aerial image using building occlusion map generated by perspective projection of digital surface model (DSM) made from ALS; 5) Filtering moving vehicles from the aerial image using multiple aerial images; 6) Extracting road markings from the aerial image utilizing adaptive thresholding; 7) Generating normal distribution map from aerial road markings which was considered as the reference for the matching; 8) Subdividing each MMS survey into small patches; 9) defining a dynamic sliding window over patches to register the target patch in the middle of the window; 10) Register the dynamic window to the aerial reference using normal distribution transform (NDT); 11) Adjust the MMS trajectory and regenerate the MMS data. This method can achieve high accuracy and precision without performing labor-intensive and costly landmark update technique which requires field survey for hundreds of GCP acquisition and manually picking the points in the point cloud. In most of the cases, the accuracy of the output generated using our proposed method can even outperform the traditional landmark updating.

The experimental results from the several surveys in one of the dense urban areas in Tokyo were employed to assess the performance of the proposed MMS registration method. In total, the proposed method could decrease the average MMS data error from 99.7cm to 12.6cm, while the maximum error was limited from more than two meters to around half-meter. The results proved that the performance of the proposed method was even better than the labor-intensive and time-consuming landmark update method, without requiring the field survey for acquiring the ground control points (GCPs) and manually picking the points in the point cloud data.

In order to make the large size point clouds of the city with billions of points applicable for a variety of applications, an automatic 3D building modeling approach is

proposed in the third part of this dissertation. The proposed method takes the combined ALS and MMS point cloud as input and generates vector based 3D building models with much less information. To generate these three-dimensional models, first the buildings segmentation is performed utilizing the 2D boundary map provided by governments. Then, a method based on the z-coordinate histogram is proposed to segment the roofs of each building. Later, the 3D shape of each roof is reconstructed by feature detection using RANSAC and the proposed roof reconstruction algorithm. Finally, reconstructed buildings are put in the corresponding height to generate the 3D model of each building. 3D building map of the Shinjuku area was reconstructed point cloud data to evaluate the proposed 3D modeling framework. The results showed that the proposed method is able to reconstruct the 3D model of the buildings in a fast and accurate manner.

## **6.2. Future work**

The proposed MMS registration method was limited to the road markings and therefore in the areas without road markings it could not achieve the required accuracy. In the future we will extend this method to more general feature matching framework which uses other features such as building edges. In this case the relief displacement of the buildings in the aerial image should be considered. Also we should consider the 2.5D characteristics of the aerial data where the MMS can acquire 3D street view information. If these differences are not taken into account, the registration quality will be significantly affected. Furthermore,

In this dissertation, the final reconstruction step was limited to the 3D building generation. In the future we will extend this for other objects such as traffic signs, traffic light, guard rails, etc. In addition, we will work on the reconstruction of the vector based road markings from point cloud to be able to use the output directly for the ADAS map.

Another future direction is applying the reconstructed maps for the localization and navigation of the intelligent vehicles and see how they can assist the precision and computation time, and finally generate the localization accuracy map which defines the quality of map in different areas for the vehicle localization. This will be obtained by simulating different map matching based localization techniques such as ICP and NDT to define the registration quality. In addition, we can consider the satellite visibility to define the quality of sensor fusion for the localization.

## **Appendix.**

# **GCL Social Project: Autonomous Vehicle Platform for Smart Society**

### **Project description**

Autonomous vehicles (AV) are an upcoming technology that redefines our society. Automakers are announcing that they are rolling out the AV technology over the next couple of years, and we have to prepare for this significant change in mobility and transportation of our cities. Although AVs can provide a significant comfort and solve or alleviate many social challenges by introducing increased safety, higher efficiency, less energy consumption, and more productivity, they will also introduce new concerns to our society. In this project, we will take steps toward designing a safe and reliable autonomous vehicle platform which enhances the possible benefits of this technology. This platform will be employed by other GCL students to implement their social innovation project and make our society a better place to live. The members of this project are Mahdi Javanmardi and Ehsan Javanmardi who worked together during their Ph.D. program as a group, from the learnings and discussions about different aspects of AVs and their social impacts to the development and implementation of the essentials of the AV such as localization, mapping, and obstacle detection frameworks. The output of this project will be partially implemented on the GCL MUSCAT (Mobile Unit for Smart Campus Transportations) platform under the supervision of professor Tsukada and accessible for all GCL students who are interested.

This project is contained four main stages: (1) learning; (2) design; (3) implementation; and (4) social reflection. The first stage, the learning, means acquiring knowledge about different aspects of AVs, such as business model, social benefits, social concerns, becoming aware of the state-of-the-art research, and touch on the promising technologies under development by the pioneer companies in this field. The next is the design, which is to understand the requirements of an AV and propose a novel and innovative framework which is able to accurately localize itself, map the surrounding environment, and detect and avoid the obstacles along its trajectory. After designing the framework, each module should be carefully implemented. In this stage, the best fitting platforms are chosen to develop and implement the framework. After the implementation, our platform is ready for the test. Since our platform is designed for solving different social problems, it should be introduced to the GCL students to boost their social innovation project and get their feedback to define the platform improvement priorities. To achieve this, a global design workshop was planned to experience and discuss a safe AV experiment design, practically.

Definitely, to be able to design and implement a platform that can be used by a variety of people in our society, the first stage which is learning the essentials, benefits, and concerns is critical. For designing a successful platform, only a high-end technology is not enough. We have to know about the government strategies, business models, investments, and development challenges. To gain this valuable knowledge, I have attended several meeting of the ITS Japan, which enjoys the full backing of the Japanese government and national agencies. In the sub-working group of “Added values of map,” I could learn the government strategies, business models, investment problems, and development challenges from the frontier companies in the field of mapping for

autonomous driving and discuss with them face-to-face. Moreover, I have intensively communicated with the big mapping companies working on the AV applications such as ATEC and AISAN TECHNOLOGY to touch on the state-of-the-art technology and learn the existing challenges. Furthermore, communicating with different key persons in companies such as ZMP, U's factory, and Tier IV increased my knowledge toward the primary goal of this project.

To improve the research and communication skills, gain a comprehensive international research experience, and work with the pioneer research groups in the autonomous vehicle field, I have visited U.C. Berkeley as a visiting research scholar for a half-year. During that period, I worked with PATH project under the supervision of professor Wei-bin Zhang to design a Lidar-based obstacle detection framework and develop a data collection platform for the AV research. Also, I worked with the Berkeley DeepDrive project to generate a dataset for the obstacle detection<sup>1</sup>, motion generation, and comprehension with social interaction<sup>2</sup>. Moreover, I had an opportunity to share ideas and learn from companies such as Honda, Valeo, and Velodyne during different meetings and discussions.

In this project, two stages of design and implementation were tightly connected. The developed framework consists of three main components: (1) mapping; (2) localization; and (3) obstacle detection. The design and implementation of the first component, the mapping, was mainly done by me during this dissertation. The second component, the localization, was designed and developed by Ehsan Javanmardi. Finally, the obstacle detection component was designed and developed by a mutual effort between

---

<sup>1</sup> <https://deepdrive.berkeley.edu/project/3d-object-detection-based-lidar-and-camera-fusion>

<sup>2</sup> <https://deepdrive.berkeley.edu/project/generic-motion-generation-and-comprehension-social-interactions>

Ehsan Javanmardi and me during the visit at U.C. Berkeley. These components will be partially implemented on the GCL MUSCAT platform and can be used in the social innovation projects.

In order to understand the impact of the AVs for the society, I had an opportunity to organize a group discussion between GCL students during the course “Practical English for GCL students” with the title of “How Autonomous Cars Will Change Our Society and Economy — Think Like a Government and Set Strategies!”. In this lecture, different groups had to put themselves in the position of the Japanese government and set strategies to highlight the benefits of the emerging AV technologies to the society, but limit the negative impacts and the concerns at the same time.

Finally, I have designed a GCL global design workshop (GDWS) titled “Safe and Effective AV experiment Design” using the MUSCAT platform. The purpose of this workshop is to strengthen the participants' knowledge about the key AV components and increase their skills in designing a safe and effective autonomous driving experiment. In the following section, details of the workshop will be introduced.

## **GCL Workshop: Safe and effective AV experiment design**

### **Concept, objectives and methodology of the workshop**

“Self-driving cars” are a promising future technology that brings significant impacts and changes to our society. A smooth and automated mobility platform can help many GCL students to boost their social projects. However, the implementation of a safe and effective autonomous driving experiment requires knowledge and experience. This workshop aims to introduce the MUSCAT platform to those students and strengthen their skills in designing a safe and effective autonomous driving experiment. By attending this workshop, attendees will become familiar with key requirements of the driverless cars

and have a self-driving experience. During the self-driving experience, they will learn the basics of the safe AV experiment design consisting the experiment planning, required check lists, safety considerations, dealing with emergency cases, and reporting. Moreover, this workshop will be an opportunity for them to think how MUSCAT can assist their projects, and provide us with feedbacks to defined development priorities.

Objectives of the workshop are as follows:

- Learn about the concept of the self-driving cars and basic localization and mapping
- Learn about possible social benefits and concerns related to AVs
- Strengthen the skills of safe and effective AV experiment design
- Experience a self-driving car in the campus
- Make a group discussion to improve the “MUSCAT safe experiment manual”
- Get feedbacks to define the future development priorities

In this workshop, we expect from students to learn basics of the autonomous driving experiment and experience the self-driving in the campus. Then, make a group work to improve the existing manual for the in-campus experiments “MUSCAT safe experiment manual.” To do this, the workshop will start with a lecture on the driverless cars. This lecture includes the basics of the driverless cars and their two key requirements: localization and mapping. After learning these concepts, the students are invited to have a real self-driving experience using MUSCAT platform. Then after a short coffee break, students are encouraged to discuss the potential requirements for a safe in-campus experiment. As a result, they should make suggestions to improve the experiment manual. Their reflections will be employed to redefine the manual and future development priorities of the MUSCAT project.

## Program of workshop

Program of the workshop is as follows:

1. Lecture on “Self-Driving Cars: Basics of Localization and Mapping”
2. Workshop introduction (Explain the overall flow of the workshop)
3. Self-driving experience using MUSCAT platform
4. Group discussion about potentials requirements for a safe in-campus experiment
5. Reflection
6. Wrap up

Time table of the workshop is as follows:

<i>Time</i>	<i>Description</i>	<i>Attendees' role</i>
10:00	Start (Self-introduction)	
10:15	Lecture on “Self-driving cars: Localization and Mappings”	Learning the concept of driverless car and how it works.
11:15	Workshop introduction	
11:30	Self-driving demonstration by instructors	Become familiar with the platform and methodology
12:00	Lunch time	
1:00 (2.5hr)	Self-driving experience by students	Experience real self-driving as a final user to be able to understand the feeling of passengers. Learn how to design a safe and effective autonomous driving experiments and what the existing challenges are.
3:30	Coffee break	
3:45	Group discussion description	
4:00 (1.5hr)	Group discussion	Discuss about how to achieve a safe and effective in-campus experiments. Provide suggestions to improve the MUSCAT experiment manual.

## Appendix. GCL Social Project: Autonomous Vehicle Platform for Smart Society

---

5:30	Reflection	Each person gives 3~5mins speech about his/her conclusions.
6:00	Wrap up	Learn about development priorities of the MUSCAT
6:30	End	

Required human resource:

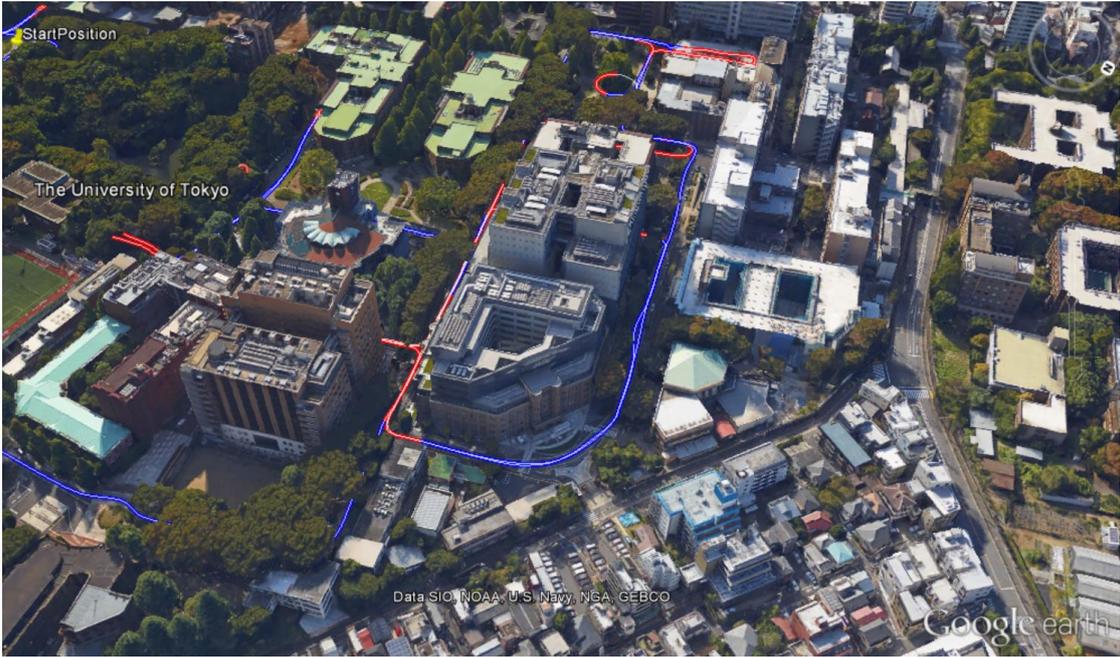
- Photographer for archive (1 person)
- Lecturer for the introduction (2 persons)
- Safety staffs for the self-driving experience (3 persons)

Other resources:

- MUSCAT (Brain computer, LiDAR, Stereo camera, and IMU)
- Microsoft Surface for visualizing the data on the MUSCAT
- Computer for the simulation of self-driving (1 system)
- Camera for archive
- Traffic cone for limiting part of the campus



Candidate place to hold the lectures and discussions: GCL Lab



Candidate routes for the AV experiment (Google earth)

## ***GCL acknowledgment***

This research was supported by Graduate Program for Social ICT Global Creative Leaders (GCL) of The University of Tokyo by Japan society for the promotion of Science (JSPS).

I would like to express my appreciation to Prof. Kamijo for giving me a unique chance to pursue the GCL program. I also want to thank Prof. Kuniyoshi, Prof. Asami, Prof. Naemura, Prof. Tsukada, Prof. Ohnishi, Prof. Aida, Prof. Gharebaghi, and all other GCL professors and staffs.

I would like to give special thanks to Prof. Wei-Bin Zhang for the opportunity of the half-year research visit at the California PATH Program of the University of California, Berkeley, and all the support he has shown me throughout my research and career. I would also like to thank all PATH members and colleagues for fruitful and insightful discussions and contributions.

Mahdi Javanmardi

ジャワンマーディ マハディ

## *List of Publications*

1. Mahdi Javanmardi, Yanlei Gu, Ehsan Javanmardi, Li-Ta Hsu, and Shunsuke Kamijo, “3D Building Map Reconstruction in Dense Urban Areas by Integrating Airborne Laser Point Cloud with 2D Boundary map,” Presented at IEEE International Conference on Vehicular Electronics and Safety (ICVES), Nov 2015.
2. Mahdi Javanmardi, Ehsan Javanmardi, Yanlei Gu, and Shunsuke Kamijo, “A Novel Approach for Post-Calibration of Mobile Mapping Systems Using Intensity Reflection and Airborne Imagery,” Presented at *Transportation Research Board 96th Annual Meeting (TRB)*, Jan 2017.
3. Mahdi Javanmardi, Ehsan Javanmardi, Yanlei Gu, and Shunsuke Kamijo, “Automatic Calibration of 3D Mobile Laser Scanning using Aerial Surveillance Data for Precise Urban Mapping,” Presented at *IEEE Intelligent Vehicles Symposium (IV)*, June 2017.
4. Mahdi Javanmardi, Ehsan Javanmardi, Yanlei Gu, and Shunsuke Kamijo, “Precise Mobile Laser Scanning for Urban Mapping Utilizing 3D Aerial Surveillance Data,” To be presented *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, October 2017.
5. Ehsan Javanmardi, Mahdi Javanmardi, Yanlei Gu, and Shunsuke Kamijo, “Autonomous Vehicle Self-Localization Based on Multilayer 2D Vector Map and LiDAR,” Presented at *IEEE Intelligent Vehicles Symposium (IV)*, June 2017.
6. Ehsan Javanmardi, Mahdi Javanmardi, Yanlei Gu, and Shunsuke Kamijo, “Autonomous Vehicle Self-Localization Based on Probabilistic Planar Surface Map and Multi-channel LiDAR in Urban Area,” To be presented at *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, October 2017.

## References

- [1] H. Wu, Z. He, and J. Gong, "A virtual globe-based 3D visualization and interactive framework for public participation in urban planning processes," *Computers, Environment and Urban Systems*, vol. 34, no. 4, pp. 291–298, Jul. 2010.
- [2] S. Hur *et al.*, "28 GHz channel modeling using 3D ray-tracing in urban environments," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, 2015, pp. 1–5.
- [3] F. Xu, X. Chen, A. Ren, and X. Lu, "Earthquake disaster simulation for an urban area, with GIS, CAD, FEA, and VR integration," *Tsinghua Science & Technology*, vol. 13, Supplement 1, pp. 311–316, Oct. 2008.
- [4] R. W. Wolcott and R. M. Eustice, "Visual localization within LIDAR maps for automated urban driving," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 176–183.
- [5] Y. Gu, L. T. Hsu, and S. Kamijo, "GNSS/onboard inertial sensor integration with the aid of 3D building map for lane-level vehicle self-localization in urban canyon," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 4274–4287, Jun. 2016.
- [6] S. Wang, S. Fidler, and R. Urtasun, "Holistic 3D scene understanding from a single geo-tagged image," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3964–3972.
- [7] N. Sairam, S. Nagarajan, and S. Ornitz, "development of mobile mapping system for 3D road asset inventory," *Sensors*, vol. 16, no. 3, p. 367, Mar. 2016.
- [8] L. Zhu and J. Hyypä, "The use of airborne and mobile laser scanning for modeling railway environments in 3D," *Remote Sensing*, vol. 6, no. 4, pp. 3075–3100, Apr. 2014.
- [9] M. Holopainen *et al.*, "Tree mapping using airborne, terrestrial and mobile laser scanning – A case study in a heterogeneous urban forest," *Urban Forestry & Urban Greening*, vol. 12, no. 4, pp. 546–553, 2013.
- [10] C. Fruh and A. Zakhor, "Constructing 3D city models by merging aerial and ground views," *IEEE Computer Graphics and Applications*, vol. 23, no. 6, pp. 52–61, Nov. 2003.
- [11] Rongxing Li, "Mobile mapping: an emerging technology for spatial data acquisition," *Journal of the American Society for Photogrammetry and Remote Sensing*, vol. 63, no. 9.
- [12] M. J. Olsen, "Guidelines for the Use of Mobile LIDAR in Transportation Applications." Transportation Research Board, 2013.
- [13] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: review and future perspectives," *IEEE Intelligent Transportation*

- Systems Magazine*, vol. 6, no. 4, pp. 6–22, winter 2014.
- [14] B. Ranft and C. Stiller, “The role of machine vision for intelligent vehicles,” *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 8–19, Mar. 2016.
- [15] J. Berssenbrügge, A. Trächtler, and C. Schmidt, “Visualization of headlight illumination for the virtual prototyping of light-based driver-assistance systems,” *Journal of Computing and Information Science in Engineering*, vol. 16, no. 3, pp. 030901–030901-7, Jun. 2016.
- [16] B.-H. Lee, J.-H. Song, J.-H. Im, S.-H. Im, M.-B. Heo, and G.-I. Jee, “GPS/DR error estimation for autonomous vehicle localization,” *Sensors*, vol. 15, no. 8, pp. 20779–20798, Aug. 2015.
- [17] Z. Jiang and P. D. Groves, “NLOS GPS signal detection using a dual-polarisation antenna,” *GPS Solution*, vol. 18, no. 1, pp. 15–26, Jan. 2014.
- [18] Y. Gu, L.-T. Hsu, and S. Kamijo, “Passive sensor integration for vehicle self-localization in urban traffic environment,” *Sensors*, vol. 15, no. 12, pp. 30199–30220, Dec. 2015.
- [19] S. Kamijo, Y. Gu, and L.-T. Hsu, “Autonomous vehicle technologies: localization and mapping,” *IEICE ESS Fundamentals Review*, vol. 9, no. 2, pp. 131–141, 2015.
- [20] J. Zhang and S. Singh, “LOAM: Lidar odometry and mapping in real-time,” in *Robotics: Science and Systems Conference (RSS 2014)*, 2014.
- [21] H. Kim, B. Liu, C. Y. Goh, S. Lee, and H. Myung, “Robust vehicle localization using entropy-weighted particle filter-based data fusion of vertical and road intensity information for a large scale urban area,” *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1518–1524, Jul. 2017.
- [22] K. Yoneda, C. Yang, S. Mita, T. Okuya, and K. Muto, “Urban road localization by using multiple layer map matching and line segment matching,” in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 525–530.
- [23] J. Choi, “Hybrid map-based SLAM using a Velodyne laser scanner,” in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, pp. 3082–3087.
- [24] I. Cvišić and I. Petrović, “Stereo odometry based on careful feature selection and tracking,” in *2015 European Conference on Mobile Robots (ECMR)*, 2015, pp. 1–6.
- [25] J. Ziegler *et al.*, “Making Bertha drive #x2014;an autonomous journey on a historic route,” *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 2, pp. 8–20, Summer 2014.
- [26] A. Broggi *et al.*, “PROUD-public road urban driverless-car test,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3508–3519, Dec. 2015.
- [27] D. Pordel, L. Petersson, and P. Kilby, “dMap: A low-cost distributed mapping concept for future road asset management,” in *Proceedings of the International Conference on Distributed Smart Cameras*, New York, NY, USA, 2014, p. 24:1–24:8.
- [28] T. E. El-Diraby, S. Kinawy, and S. M. Piryonisi, “A comprehensive review of approaches used by Ontario municipalities to develop road asset management plans,” presented at the

- 
- Transportation Research Board 96th Annual Meeting Transportation Research Board*, 2017.
- [29] H. C. Nguyen, G. R. MacCartney, T. Thomas, T. S. Rappaport, B. Vejlgard, and P. Mogensen, "Evaluation of empirical ray-tracing model for an urban outdoor scenario at 73 GHz e-band," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, 2014, pp. 1–6.
- [30] G. R. Maccartney and T. S. Rappaport, "73 GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York City," presented at the 2014 1st *IEEE International Conference on Communications, ICC 2014*, 2014.
- [31] Y. Kishiki *et al.*, "Implementation of reflection on curved surfaces and physical optics in ray tracing for tunnel propagation," *IEICE Transactions on Electronics*, vol. E96–C, no. 1, pp. 42–50, Jan. 2013.
- [32] A. Kumar and T. R. Rao, "Analysis of planning and deployment issues for short range Gigabit radio's at 60 GHz," in *2013 International Conference on Communication and Signal Processing*, 2013, pp. 24–28.
- [33] A. Dhekne, M. Gowda, and R. R. Choudhury, "Cell tower extension through drones: poster," in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, New York, NY, USA, 2016, pp. 456–457.
- [34] M. Batty, "Urban modeling," in *International Encyclopedia of Human Geography*, R. Kitchin and N. Thrift, Eds. Oxford: Elsevier, 2009, pp. 51–58.
- [35] G. Herbert and X. Chen, "A comparison of usefulness of 2D and 3D representations of urban planning," *Cartography and Geographic Information Science*, vol. 42, no. 1, pp. 22–32, Jan. 2015.
- [36] L. Yin, "Integrating 3D visualization and GIS in planning education," *Journal of Geography in Higher Education*, vol. 34, no. 3, pp. 419–438, Aug. 2010.
- [37] M. Kristan *et al.*, "Rapid 3D City model approximation from publicly available geographic data sources and georeferenced aerial images," presented at the *17th Computer Vision Winter Workshop*, Mala Nedelja, Slovenia, 2012.
- [38] F. Leberl, H. Bischof, T. Pock, A. Irschara, and S. Kluckner, "Aerial Computer vision for a 3D virtual habitat," *Computer*, vol. 43, no. 6, pp. 24–31, Jun. 2010.
- [39] B. Rodríguez-Cuenca, S. García-Cortés, C. Ordóñez, and M. C. Alonso, "Automatic detection and classification of pole-like objects in urban point cloud data using an anomaly detection algorithm," *Remote Sensing*, vol. 7, no. 10, pp. 12680–12703, Sep. 2015.
- [40] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, winter 2014.
- [41] G. P. Gwon, W. S. Hur, S. W. Kim, and S. W. Seo, "Generation of a precise and efficient lane-level road map for intelligent vehicle systems," *IEEE Transactions on Vehicular Technology*,
-

- vol. PP, no. 99, pp. 1–1, 2016.
- [42] L. Cheng, Y. Wu, L. Tong, Y. Chen, and M. Li, “Hierarchical registration method for airborne and vehicle Lidar point cloud,” *Remote Sensing*, vol. 7, no. 10, pp. 13921–13944, Oct. 2015.
- [43] M. Weinmann, M. Weinmann, C. Mallet, and M. Brédif, “A classification-segmentation framework for the detection of individual trees in dense mms point cloud data acquired in urban areas,” *Remote Sensing*, vol. 9, no. 3, p. 277, Mar. 2017.
- [44] F. Monnier, B. Vallet, and B. Soheilian, “Trees detection from laser point clouds acquired in dense urban areas by a mobile mapping system,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 245–250, Jul. 2012.
- [45] B. Ranft and C. Stiller, “The role of machine vision for intelligent vehicles,” *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 8–19, Mar. 2016.
- [46] R. W. Wolcott and R. M. Eustice, “Robust LIDAR localization using multiresolution Gaussian mixture maps for autonomous driving,” *The Int’l Journal of Robotics Research*, vol. 36, no. 3, pp. 292–319, Mar. 2017.
- [47] H. Jin, M. Miska, E. Chung, M. Li, and Y. Feng, “Road feature extraction from high resolution aerial images upon rural regions based on multi-resolution image analysis and Gabor filters,” 2012.
- [48] C. Ellul, M. Adjrad, and P. Groves, “The impact of 3d data quality on improving GNSS performance using city models initial simulations,” in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, vol. IV-2-W1, pp. 171–178.
- [49] R. W. Wolcott and R. M. Eustice, “Visual localization within LIDAR maps for automated urban driving,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 176–183.
- [50] H. Kaartinen, J. Hyyppä, A. Kukko, A. Jaakkola, and H. Hyyppä, “Benchmarking the Performance of Mobile Laser Scanning Systems Using a Permanent Test Field,” *Sensors*, vol. 12, no. 9, pp. 12814–12835, Sep. 2012.
- [51] K. Ishikawa, J. Takiguchi, Y. Amano, and T. Hashizume, “A Mobile Mapping System for road data capture based on 3D road model,” in *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications*, 2006, pp. 638–643.
- [52] N. Kajiwara, J. Takiguti, R. Hirokawa, and Y. Shima, “Landmark update of GPS/INS for mobile mapping system using a nearly horizontal single camera and 3D point-cloud,” in *International Symposium on GPS/GNSS*, 2008.
- [53] Higinio González-Jorge, Joaquín Martínez Sánchez, Lucía Díaz-Vilariño, Iván Puente, and Pedro Arias, “Automatic Registration of mobile LiDAR data using high-reflectivity traffic signs,” *Journal of Construction Engineering and Management*, vol. 142, no. 8, Aug. 2016.

- 
- [54] L. Yao *et al.*, “Registration of vehicle-borne point clouds and panoramic images based on sensor constellations,” *Sensors*, vol. 17, no. 4, p. 837, Apr. 2017.
- [55] I. A and V. D, “Sensor fusion: generating 3D by combining airborne and tripod-mounted LIDAR data,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2003, vol. XXXIV-5/W10.
- [56] L. Cheng, L. Tong, Y. Wu, Y. Chen, and M. Li, “Shiftable leading point method for high accuracy registration of airborne and terrestrial LiDAR data,” *Remote Sensing*, vol. 7, no. 2, pp. 1915–1936, Feb. 2015.
- [57] B. Yang, Y. Zang, Z. Dong, and R. Huang, “An automated method to register airborne and terrestrial laser scanning point clouds,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 109, pp. 62–76, Nov. 2015.
- [58] M. Kedzierski and A. Fryskowska, “Terrestrial and aerial laser scanning data integration using wavelet analysis for the purpose of 3D building modeling,” *Sensors*, vol. 14, no. 7, pp. 12070–12092, Jul. 2014.
- [59] T.-A. Teo and S.-H. Huang, “Surface-based registration of airborne and terrestrial mobile lidar point clouds,” *Remote Sensing*, vol. 6, no. 12, pp. 12686–12707, Dec. 2014.
- [60] L. Cheng, L. Tong, M. Li, and Y. Liu, “Semi-automatic registration of airborne and terrestrial laser scanning data using building corner matching with boundaries as reliability check,” *Remote Sensing*, vol. 5, no. 12, pp. 6260–6283, Nov. 2013.
- [61] P. Polewski, A. Erickson, W. Yao, N. Coops, P. Krzystek, and U. Stilla, “Object-based coregistration of terrestrial photogrammetric and ALS point clouds in forested areas,” in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, vol. III-3, pp. 347–354.
- [62] R. Kümmerle, B. Steder, C. Dornhege, A. Kleiner, G. Grisetti, and W. Burgard, “Large scale graph-based SLAM using aerial images as prior information,” *Auton Robot*, vol. 30, no. 1, pp. 25–39, Jan. 2011.
- [63] A. Gruen *et al.*, “Joint processing of UAV imagery and terrestrial MMS data for very high resolution 3D city modeling,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W2, no. 1, 2013.
- [64] O. Tournaire, B. Soheilian, and N. Paparoditis, “Towards a sub-decimetric georeferencing of ground-based mobile mapping systems in urban areas: matching ground-based and aerial-based imagery using roadmarks - semantic scholar,” in *ISPRS Commission I Symposium*, Marne-la-Vallée, France, 2006.
- [65] Z. Hussnain, S. Oude Elberink, and G. Vosselman, “Automatic feature detection, description and matching from mobile laser scanning data and aerial imagery,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-
-

- B1, pp. 609–616, Jun. 2016.
- [66] P. Jende, M. Peter, M. Gerke, and G. Vosselman, “Advanced tie feature matching for the registration of mobile mapping imaging data and aerial imagery,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B1, pp. 617–623, Jun. 2016.
- [67] P. Jende, Z. Hussnain, M. Peter, S. Oude Elberink, M. Gerke, and G. Vosselman, “Low-level tie feature extraction of mobile mapping data (MLS/images) and aerial imagery,” in *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, vol. XL-3-W4, pp. 19–26.
- [68] M. Javanmardi, E. Javanmardi, Y. Gu, and S. Kamijo, “A novel approach for post-calibration of mobile mapping systems using intensity reflection and airborne imagery,” presented at the *Transportation Research Board 96th Annual Meeting Transportation Research Board*, 2017.
- [69] Y. Li, L. Li, D. Li, F. Yang, and Y. Liu, “A density-based clustering method for urban scene mobile laser scanning data segmentation,” *Remote Sensing*, vol. 9, no. 4, p. 331, Mar. 2017.
- [70] X. Meng, N. Currit, and K. Zhao, “Ground filtering algorithms for airborne lidar data: a review of critical issues,” *Remote Sensing*, vol. 2, no. 3, pp. 833–860, Mar. 2010.
- [71] B. Yang, R. Huang, Z. Dong, Y. Zang, and J. Li, “Two-step adaptive extraction method for ground points and breaklines from lidar point clouds,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 119, pp. 373–389, Sep. 2016.
- [72] H. Guan, J. Li, Y. Yu, Z. Ji, and C. Wang, “Using mobile LiDAR data for rapidly updating road markings,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2457–2466, Oct. 2015.
- [73] W. Zhang *et al.*, “An easy-to-use airborne LiDAR data filtering method based on cloth simulation,” *Remote Sensing*, vol. 8, no. 6, p. 501, Jun. 2016.
- [74] S. Kaasalainen, A. Jaakkola, M. Kaasalainen, A. Krooks, and A. Kukko, “Analysis of incidence angle and distance effects on terrestrial laser scanner intensity: search for correction methods,” *Remote Sensing*, vol. 3, no. 10, pp. 2207–2221, Oct. 2011.
- [75] A. Kukko, S. Kaasalainen, and P. Litkey, “Effect of incidence angle on laser scanner intensity and surface data,” *Applied Optics*, vol. 47, no. 7, pp. 986–992, Mar. 2008.
- [76] A. G. Kashani, M. J. Olsen, C. E. Parrish, and N. Wilson, “A review of lidar radiometric processing: from Ad Hoc intensity correction to rigorous radiometric calibration,” *Sensors*, vol. 15, no. 11, pp. 28099–28128, Nov. 2015.
- [77] B. Jutzi and H. Gross, “Normalization of Lidar intensity data based on range and surface incidence angle,” in *In Proceedings of Laserscanning 09*, pp. 213–218.
- [78] S. Soudarissanane, R. Lindenbergh, M. Menenti, and P. Teunissen, “Incidence angle influence on the quality of terrestrial laser scanning points,” in *In Proceedings of Laserscanning 2009*,

- pp. 83–88.
- [79] M. Magnusson, N. Vaskevicius, T. Stoyanov, K. Pathak, and A. Birk, “Beyond points: Evaluating recent 3D scan-matching algorithms,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 3631–3637.
- [80] M. Magnusson, “The three-dimensional normal-distributions transform --- an efficient representation for registration, surface analysis, and loop detection,” *ResearchGate*, Dec. 2009.
- [81] P. Biber and W. Strasser, “The normal distributions transform: a new approach to laser scan matching,” in *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings*, 2003, vol. 3, pp. 2743–2748 vol.3.
- [82] T. Kaminade, T. Takubo, Y. Mae, and T. Arai, “The generation of environmental map based on a NDT grid mapping -Proposal of convergence calculation corresponding to high resolution grid-,” in *2008 IEEE International Conference on Robotics and Automation*, 2008, pp. 1874–1879.
- [83] Mitsunobu Yoshida and Mamoru Yoshida, “The latest trends in mobile mapping system,” *Mitsubishi Electric Advance Magazine*, vol. 148, pp. 2–5, Dec. 2014.
- [84] “Mobile lidar mapping for urban data capture,” in *Proc. 14th Int. Conf. Virtual Syst. Multimedia*, 2008, pp. 95–100.
- [85] S. Miura, L.-T. Hsu, F. Chen, and S. Kamijo, “GPS error correction with pseudorange evaluation using three-dimensional maps,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3104–3115, 2015.
- [86] P. Jende, F. Nex, M. Gerke, and G. Vosselman, “Fully automatic feature-based registration of mobile mapping and aerial nadir images for enabling the adjustment of mobile platform locations in GNSS-denied urban environments,” in *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017, vol. XLII-1-W1, pp. 317–323.
- [87] T. Luettel, M. Himmelsbach, and H. J. Wuensche, “Autonomous ground vehicles #x2014; concepts and a path to the future,” *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1831–1839, May 2012.
- [88] J. Levinson and S. Thrun, “Robust vehicle localization in urban environments using probabilistic maps,” in *2010 IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 4372–4378.
- [89] Y. Gu, Y. Wada, L. Hsu, and S. Kamijo, “Vehicle self-localization in urban canyon using 3D map based GPS positioning and vehicle sensors,” in *2014 International Conference on Connected Vehicles and Expo (ICCVEx)*, 2014, pp. 792–798.
- [90] M. Obst, S. Bauer, and G. Wanielik, “Urban multipath detection and mitigation with dynamic 3D maps for reliable land vehicle localization,” in *Position Location and Navigation*

- 
- Symposium (PLANS), 2012 IEEE/ION*, 2012, pp. 685–691.
- [91] L. Wang, P. D. Groves, and M. K. Ziebart, “GNSS shadow matching: improving urban positioning accuracy using a 3D city model with optimized visibility scoring scheme: GNSS shadow matching: optimized scoring scheme,” *Navigation*, vol. 60, no. 3, pp. 195–207, Sep. 2013.
- [92] S. Peyraud *et al.*, “About non-line-of-sight satellite detection and exclusion in a 3D map-aided localization algorithm,” *Sensors*, vol. 13, no. 1, pp. 829–847, Jan. 2013.
- [93] S. Miura, S. Hisaka, and S. Kamijo, “GPS multipath detection and rectification using 3D maps,” in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, pp. 1528–1534.
- [94] L.-T. Hsu, Y. Gu, and S. Kamijo, “3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation,” *GPS Solut*, pp. 1–16, Mar. 2015.
- [95] K. Konolige and M. Agrawal, “FrameSLAM: from bundle adjustment to real-time visual mapping,” *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1066–1077, Oct. 2008.
- [96] M. Meilland, A. I. Comport, and P. Rives, “Dense omnidirectional RGB-D mapping of large-scale outdoor environments for real-time localization and autonomous navigation,” *J. Field Robotics*, vol. 32, no. 4, pp. 474–503, Jun. 2015.
- [97] K. Yoneda, H. Tehrani, T. Ogawa, N. Hukuyama, and S. Mita, “Lidar scan feature for localization with highly precise 3-D map,” in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 1345–1350.
- [98] X. Yu and J. Feng, “Research on radio propagation modeling of city impact in 3D perspective,” in *2013 International Conference on Wireless Communications Signal Processing (WCSP)*, 2013, pp. 1–5.
- [99] H. Reza Ranjbar, A. Reza Gharagozlou, and A. Reza Vafaei Nejad, “3D analysis and investigation of traffic noise impact from Hemmat highway located in Tehran on buildings and surrounding areas,” *Journal of Geographic Information System*, vol. 04, no. 04, pp. 322–334, 2012.
- [100] F. Biljecki, J. Stoter, H. Ledoux, S. Zlatanova, and A. Çöltekin, “Applications of 3D city models: state of the art review,” *ISPRS International Journal of Geo-Information*, vol. 4, no. 4, pp. 2842–2889, Dec. 2015.
- [101] B. Sirmacek and R. Lindenbergh, “Active shapes for automatic 3D modeling of buildings,” *Journal of Imaging*, vol. 1, no. 1, pp. 156–179, Nov. 2015.
- [102] L. M. Martin Kada, “3D building reconstruction from LiDAR based on a cell decomposition approach,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, 2009.
- [103] N. Haala and M. Kada, “An update on automatic 3D building reconstruction,” *ISPRS Journal*
-

- of Photogrammetry and Remote Sensing*, vol. 65, no. 6, pp. 570–580, Nov. 2010.
- [104] W. Zhang, Y. Chen, K. Yan, G. Yan, and G. Zhou, “Primitive-based 3d building reconstruction method tested by reference airborne data,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIX-B3, pp. 373–378, Jul. 2012.
- [105] E. Orthuber and J. Avbelj, “3D building reconstruction from lidar point clouds by adaptive dual contouring,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3/W4, pp. 157–164, Mar. 2015.
- [106] F. Rottensteiner *et al.*, “The ISPRS benchmark on urban object classification and 3D building reconstruction,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 293–298, Jul. 2012.
- [107] R. Wang, “3D building modeling using images and LiDAR: a review,” *International Journal of Image and Data Fusion*, vol. 4, no. 4, pp. 273–292, Dec. 2013.
- [108] Y. Gu, Y. Wada, L. T. Hsu, and S. Kamijo, “SLAM with 3Dimensional-GNSS,” in *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2016, pp. 190–197.

