

Article

Towards High-Definition 3D Urban Mapping: Road Feature-Based Registration of Mobile Mapping Systems and Aerial Imagery

Mahdi Javanmardi ^{1,*}, Ehsan Javanmardi ¹, Yanlei Gu ²  and Shunsuke Kamijo ²

¹ Department of Information & Communication Engineering, the University of Tokyo, Tokyo 153-8505, Japan; ehsan@kmj.iis.u-tokyo.ac.jp

² The Institute of Industrial Science (IIS), The University of Tokyo, Tokyo 153-8505, Japan; guyanlei@kmj.iis.u-tokyo.ac.jp (Y.G.); kamijo@iis.u-tokyo.ac.jp (S.K.)

* Correspondence: mahdi@kmj.iis.u-tokyo.ac.jp; Tel.: +81-3-5452-6272

Received: 24 July 2017; Accepted: 18 September 2017; Published: 21 September 2017

Abstract: Various applications have 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 is highly dependent on the performance of the vehicle's self-localization, which is generally performed by high-end global navigation satellite system (GNSS)/inertial measurement unit (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 improve MMS accuracy by measuring ground control points (GCPs) and manually identifying those points in the data, are both labor-intensive and time-consuming. In this paper, we propose a novel and comprehensive framework for automatically georeferencing MMS data by capitalizing on road features extracted from high-resolution aerial surveillance data. 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.

Keywords: mobile mapping system; airborne imagery; airborne laser scanning; point cloud; lidar; urban; road marking; 3D map; landmark update

1. Introduction

With their ability to obtain high-definition information about 3D environments, various applications have used a mobile mapping system (MMS) as the main urban remote sensing platform. Such applications include 3D map reconstruction for intelligent vehicle navigation and control [1–4], 3D city modeling [5–7], city visualizations [8], road asset inventories [9], railway modeling [10], vegetation detection and urban forest inventories [11–13]. As the primary provider of static environment information for the intelligent vehicles, maps constructed through MMSs have even been considered as “virtual sensors” [14] that would enable long-distance autonomous navigation [15]. Advanced driver assistance systems (ADAS), which protect drivers and passengers through active, integrated safety, require maps to include precise sub-road details, such as the position and topology of pavement markings and stop lines, which are utilized for lane change assistance and lane departure warnings [16,17].

Vehicle self-localization is another application that is highly dependent on map accuracy and precision. The main vehicle-based positioning technologies, such as global navigation satellite systems (GNSSs), 3D light detection and ranging (lidar), and vision-based systems, can be assisted by access to

a prior map. A 3D building map helps GNSSs address the signal blockage and reflection problems caused by tall buildings in urban areas [18]. Another well studied positioning technique for urban area is map matching based on lidar [19,20] and vision sensors [21]. This method can achieve high positional accuracy relative to the prior map, but are dependent on map precision and accuracy for global accuracy.

Here, an important question arises regarding how high-accuracy and high-precision maps should be obtained in a time and cost-efficient way in dense urban areas with tall buildings and street canyons when no prior map exists. An MMS is a valuable geo-information acquisition technology that is specifically designed to provide fast and high-definition mapping of dense urban areas. However, its ability to acquire high-precision 3D data is constrained by the vehicle's positioning system. The position and orientation system (POS) of an MMS depends on a tight integration between a high-end GNSS and an inertial measurement unit (IMU) [22,23]. The 3D measurements acquired by the vehicle-mounted laser scanners are georeferenced to the global coordinate system based on the POS data [24]. In urban areas where the accuracy of GNSS/IMU degrades significantly, the quality of the collected data is also affected.

The map required for many of the applications listed above should be both "accurate" and "precise". Figure 1 illustrates common definitions of accuracy and precision, the two principal terms used in this type of mapping. Generally, the map is reconstructed by georeferencing multiple MMS surveys. In urban areas, each MMS survey involves an independent error, and therefore the resulting map would be neither accurate nor precise if no correction were performed (Figure 1a). In such cases, the generated map cannot be used for any of the applications mentioned above. Instead, if corrections were made, but the corrections were not able to limit the independent error from each survey to a certain threshold, the resulting map would indeed be more accurate, but still not precise (Figure 1b). Maps after applying the landmark update, which employs the reference ground control points (GCPs) acquired by a total station survey to update the MMS data, with only a limited number of GCPs fit in this category. In contrast, if the surveys are precisely registered, but global georeferencing were not performed, the final map would be precise but not accurate (Figure 1c). Methods that use simultaneous localization and mapping (SLAM), but without a global reference, fall into this category. In contrast, in ideal conditions, if all points of each survey are georeferenced precisely to their correct global positions, the resulting map would be both accurate and precise (Figure 1d). When a map is not globally accurate, data obtained from the map cannot be fused or integrated with global information, such as GNSS data. On the other hand, if the map is not precise, acquired data such as a vehicle position will be imprecise as well. Throughout the article, the terms accuracy and precision are used according to the above definitions.

Numerous studies have addressed MMS registration to improve the global accuracy of the obtained data. One standard method is the landmark update [10,25]. 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 vehicle trajectory is updated by correcting the previously estimated vehicle positions using the PCVs. Finally, the MMS data are regenerated with the new trajectory. Although this method can preserve some level of accuracy, the data precision is limited by the number of GCPs. Unfortunately, acquiring GCPs 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. [26] 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. [27] 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.

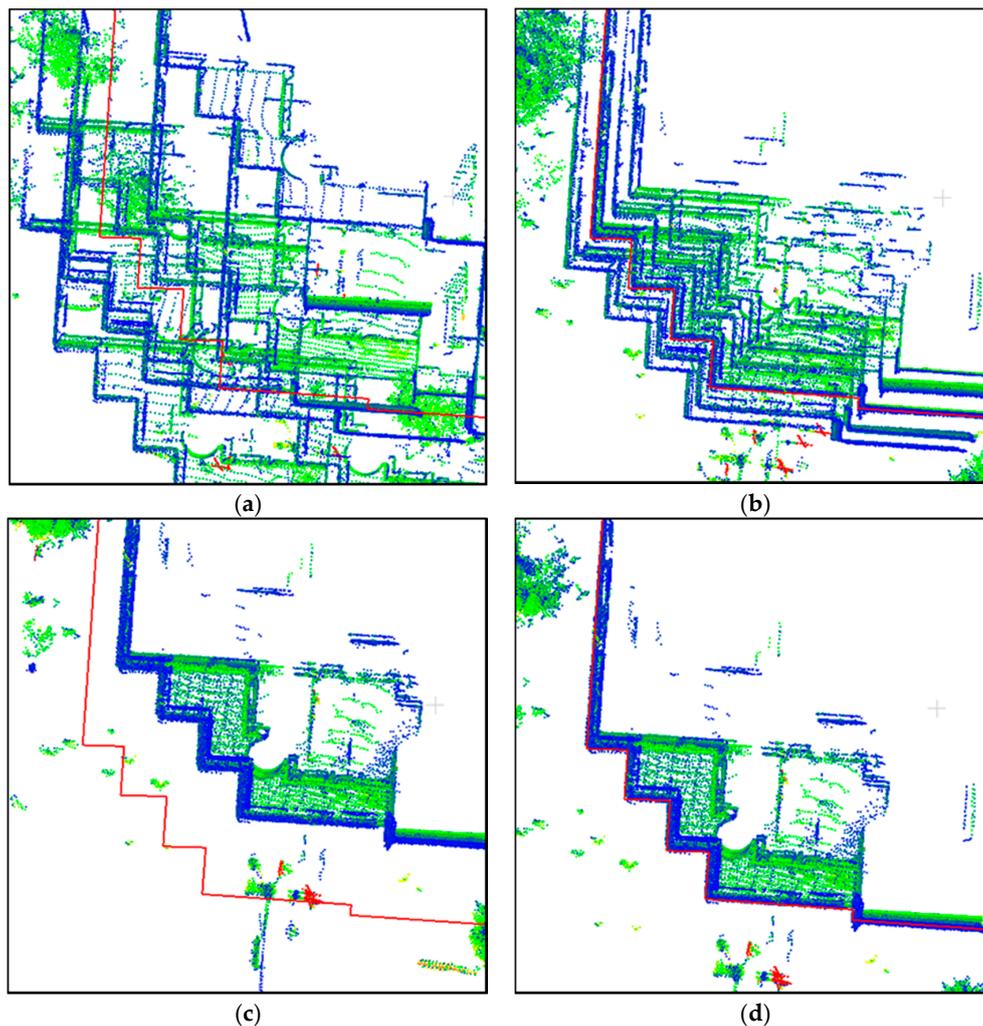


Figure 1. Visualized examples of the effects of the terms “accuracy” and “precision” in mapping (the red polyline is the ground-truth building footprint): (a) neither accurate nor precise; (b) accurate only; (c) precise only; and (d) both accurate and precise.

Aerial and ground surveillance data are two complementary data sources for high-definition 3D urban modeling [28]. Previous studies have shown that aerial surveillance data are very advantageous in maintaining the quality of ground surveys. Cheng et al. [5] 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 [29–33]. 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. [34] 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. [35] inserted the correspondences found between the range data and the aerial images as constraints into a graph-based formulation for SLAM.

Gruen [36] utilized 169 manually extracted GCPs from the geo-referenced UAV images to update the MMS data in a 2.2 km² area. Tournaire et al. [37] proposed an image-based strategy for georeferencing 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. [38] 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. [39,40] 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 [41] 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 study, 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 georeferencing 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 georeferencing the MMS surveys with a mean error of 11.6 cm.

2. Methods

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 georeferencing 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 2 illustrates the overall flow of the proposed framework.

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 [17]. 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 3a–h). Furthermore, road obstruction by tall buildings due to the relief displacement is yet another challenge. Therefore, it is necessary to develop an automatic and thorough system that more correctly extracts road marking features for the registration to address the effects described above.

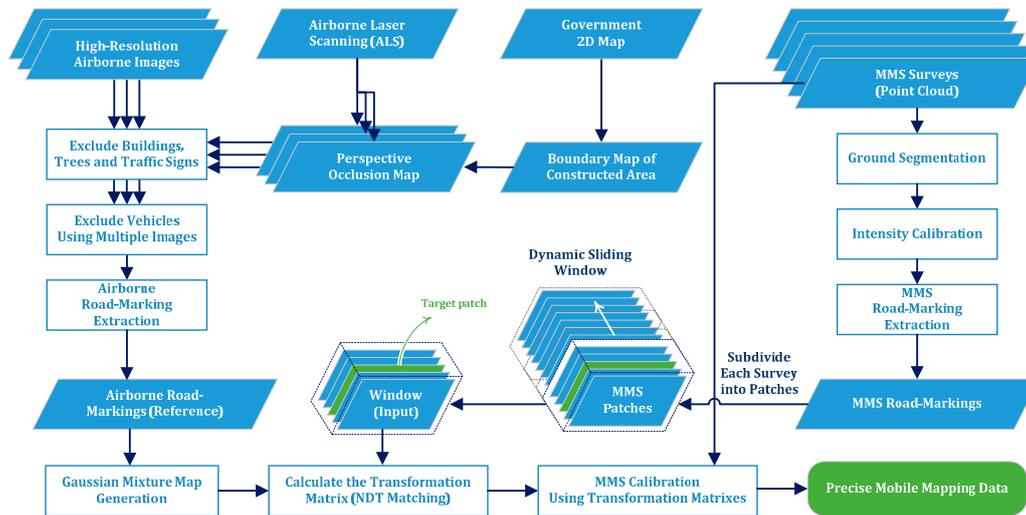


Figure 2. The overall workflow of the proposed framework.

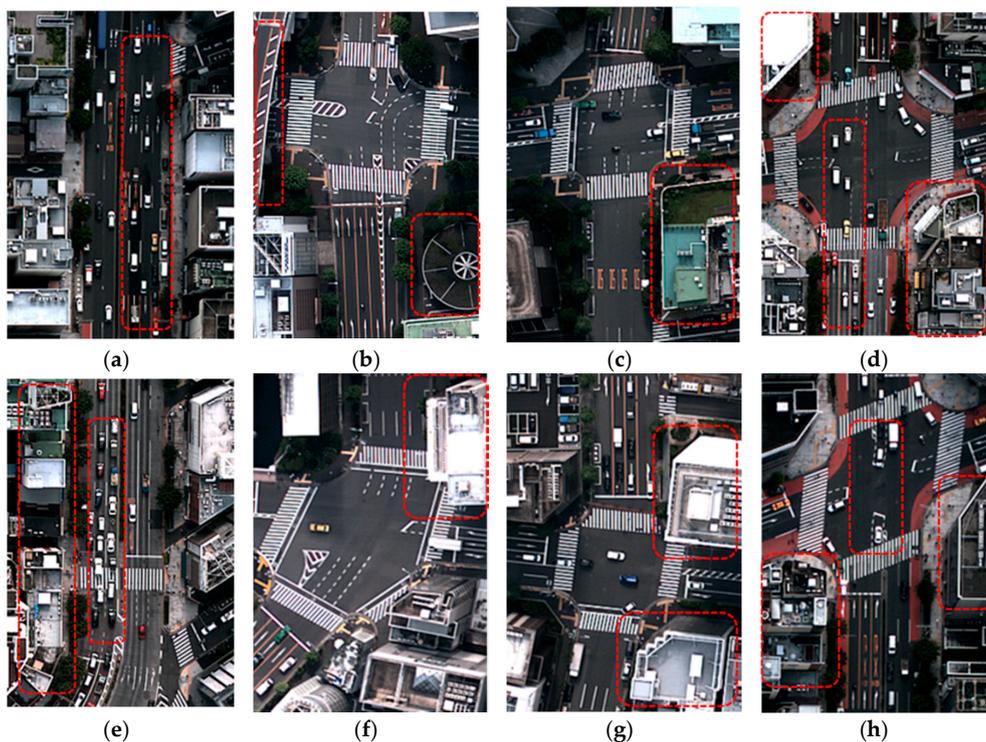


Figure 3. Sample road markings in the aerial images of the Hitotsubashi area in central Tokyo. Some of the problematic areas are highlighted by the red dotted rectangles.

In this paper, the input aerial images are calibrated, distortion-free, and georeferenced by company as a product. 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.

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 nadir 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 4). Then, the points higher than a certain threshold are considered when generating the POM. This threshold should be above ground height to ignore ground points and vehicles on the road surface when generating the POM.

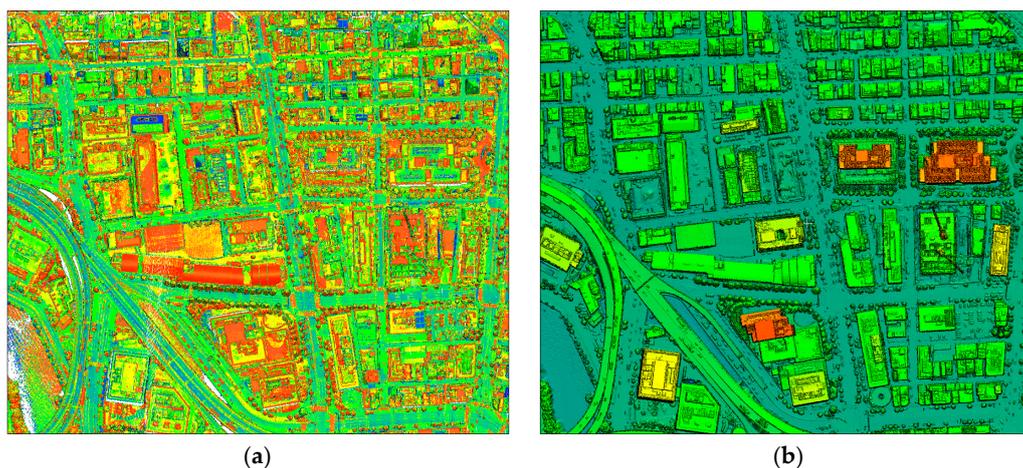


Figure 4. DSM reconstruction from ALS by 2.5D Delaunay triangulation: (a) input ALS point cloud (color denotes intensity); and (b) reconstructed DSM mesh (color denotes height).

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 5a). The resolution of the POM is defined as being equal to that of the aerial image. The aircraft's position $(c_{x,y,z})$ and orientation (ω, ϕ, κ) , 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:

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

where $c_{x,y,z}$ represents the 3D position of the camera installed on the aircraft. The transformed vertex is then projected onto the 2D image using the following equations:

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

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

where f is the focal length and (x_0, y_0) indicates the image coordinates of the principal point. Figure 5c illustrates the generated POM. To obtain clearer road segments, we also used a two-dimensional boundary map of the constructed areas that is publicly available from the Geospatial Information Authority of Japan (GSI). The final result of the road segmentation operation is shown in Figure 5d, where the generated map filters all the buildings and their relief displacements effectively.

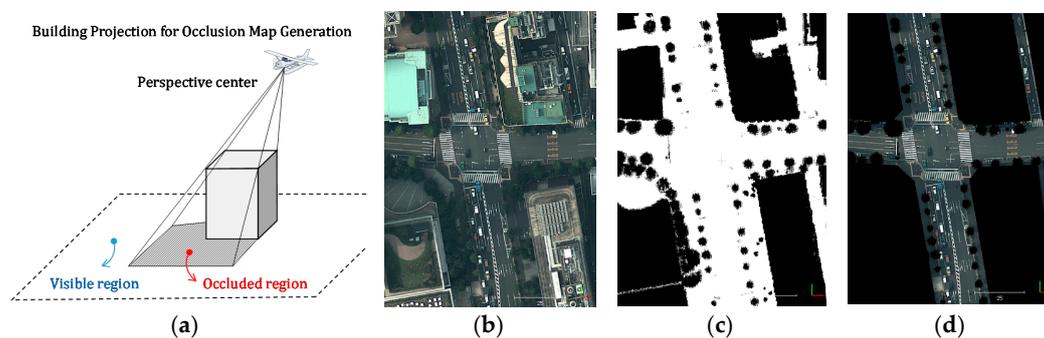


Figure 5. Generation of POM from the DSM: (a) overview of the buildings' perspective occlusion map generation; (b) original aerial image; (c) generated POM; and (d) filtered result of the aerial image.

2.1.2. Filtering Moving Vehicles from the Roads

When performing road marking extraction from aerial images, vehicles painted in light colors can mistakenly be considered as road markings. Because the objective of road marking extraction is to perform registration between ground and aerial data, the extraction does not need to be perfect. Instead, the extracted road markings are considered as features for the registration operation. Therefore, in areas with few road markings, missing a lane marking or mis-detecting a vehicle as a road marking may have a significant effect on the registration result. In contrast, in areas with sufficient features, such as intersections with zebra crossings, missing a few features will not affect the matching quality. To obtain the best vehicle filtering result, machine learning methods might be ideal; however, we did not have access to a sufficiently large and labeled data set to perform training appropriately. Therefore, we applied a simpler method that uses overlapping aerial images to filter the moving vehicles. If a vehicle is detected in the first image and it is moving, the probability that the same vehicle will appear in the second image is low, and so on. The proposed solution takes two overlapping images and compares the total of all RGB channel values for each pixel, then selects the color with a lower luminance. Vehicles in the aerial images may appear between lane markings or on zebra crossings. This simple method can not only filter all the vehicles that appear between lane markings but also all the light-colored vehicles. However, parked vehicles cannot be filtered. Moreover, dark-colored vehicles can obscure a portion of zebra crossings. Although this simple solution is imperfect, it is sufficiently robust for our application. Figure 6 illustrates these steps.

2.1.3. Extracting Road Markings from Filtered Aerial Images

The road markings extracted from the aerial image are considered as references to be used in the georeferencing of the MMS data. To obtain additional features for the registration, all types of road markings and even signs on sidewalks were considered based on the intensity contrast between the

markings and the background road surface. After a level adjustment of the image, adaptive Gaussian thresholding was performed to extract the road markings. Figure 7 illustrates an example of road markings extracted from an aerial image in bitmap format; the white pixels are the extracted road markings and the black pixels are the background.

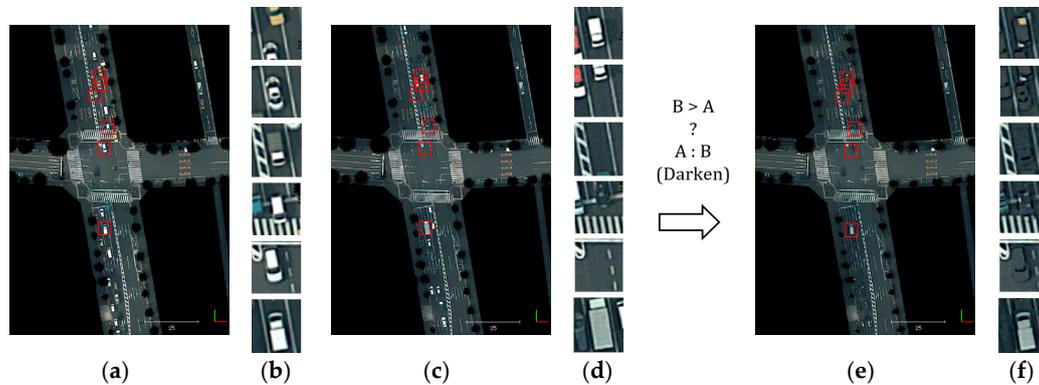


Figure 6. Filtering moving vehicles from aerial image: (a) first aerial image; (b) enlarged view of some vehicles in the first image; (c) second aerial image; (d) enlarged view of the same areas in the second image; (e) vehicle filtering result; and (f) enlarged view of the same areas in the filtered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

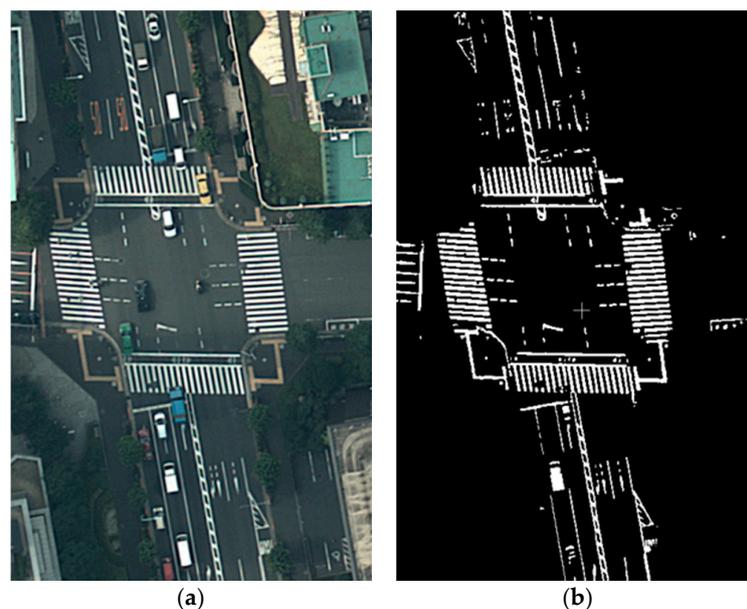


Figure 7. Road marking extraction from an aerial image by adaptive thresholding: (a) a part of the original aerial image; and (b) result of the road marking extraction process.

2.2. 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 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.

2.2.1. Ground Segmentation of the MMS Point Cloud

Ground segmentation from ground and airborne laser point cloud is a well-researched topic [10,42–44]. Curb-based road surface segmentation is a popular method for ground point extraction from MMS data [45]. 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 [46] is used for ground segmentation. Figure 8 shows the result of ground segmentation in a part of MMS point cloud.

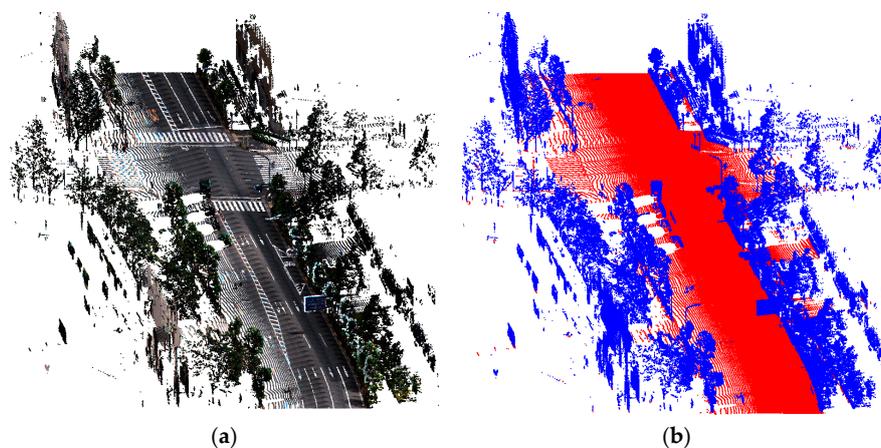


Figure 8. Road segmentation from an MMS point cloud: (a) original MMS point cloud consisting of buildings, trees, vehicles and road signs (RGB color is derived from the camera); and (b) the result of ground segmentation. The red points represent the ground, and the blue points are off-ground.

2.2.2. 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 [45]. However, the reflected laser signal is significantly affected by the scanning geometry, mainly the distance and the laser incidence angle to the target surface [47–50]; 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 [47]. 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 [51]. Although the effects of distance and incidence angle on laser intensity would ideally be modeled individually, different laser scanners from different manufacturers can show completely different responses [49]. 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 paper, the intensity value is used to differentiate the painted areas of roads and sidewalks. One road surface characteristic is flatness. As shown in Figure 9, when flat surfaces such as roads and sidewalks are considered, the incidence angle of the laser beam decreases as the range increases. In this case, we can consider the intensity changes of materials based only on the distance of the measured point from the MMS scanner, which inherently includes the effect of the incidence angle.

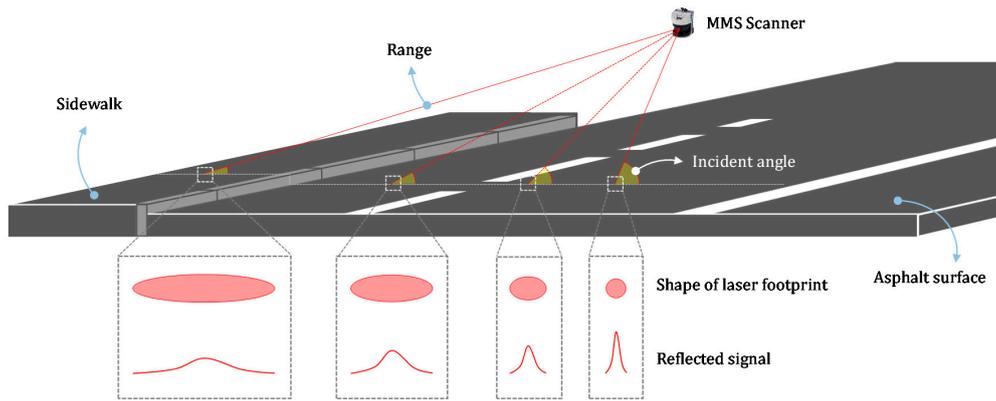


Figure 9. The effect of range and incidence angle on the shape of the laser footprint and the power of the reflected signal. As can be seen, as the range increase and the incidence angle decreases, area of the laser footprint increases, which degrades the quality of sampling.

The main idea of the intensity calibration is to normalize the intensity of both road markings and asphalt surfaces based on the intensity characteristic of the asphalt. In that case, markings on the pavement will have higher intensity compared to the asphalt background. Figure 10 shows the effect of the distance (consisting the effect of incidence angle inherently) on the reflected intensity of the asphalt surface of an experimental road using the laser installed on the MMS. In the proposed method for the calibration of these two effects, the nonlinear 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 the 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 calibration intensities, respectively; 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 11.

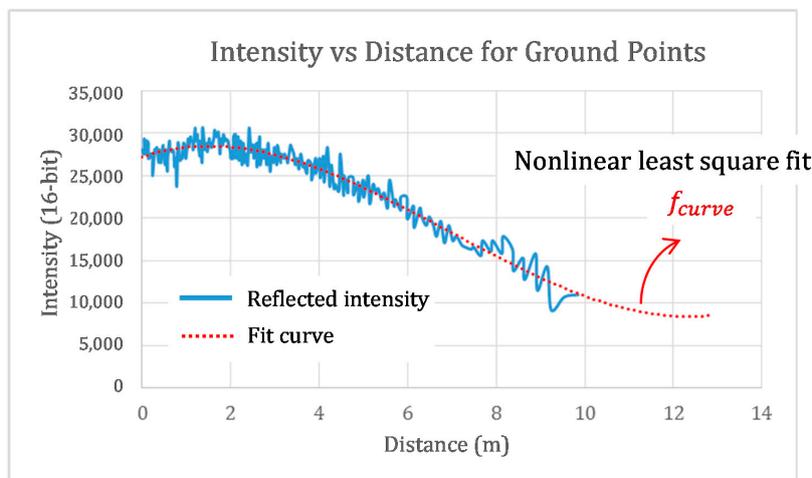


Figure 10. Effect of the increase of the distance to the reflected intensity value of the asphalt surface. By increasing distance, the angle of incidence inherently decreases. The red curve illustrates the curve fit to the relation between distance and intensity using a nonlinear least squares fitting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

As shown in Figure 11, the proposed method works well for horizontal road surfaces, but the intensity of the curbs and low vegetation can be overestimated.

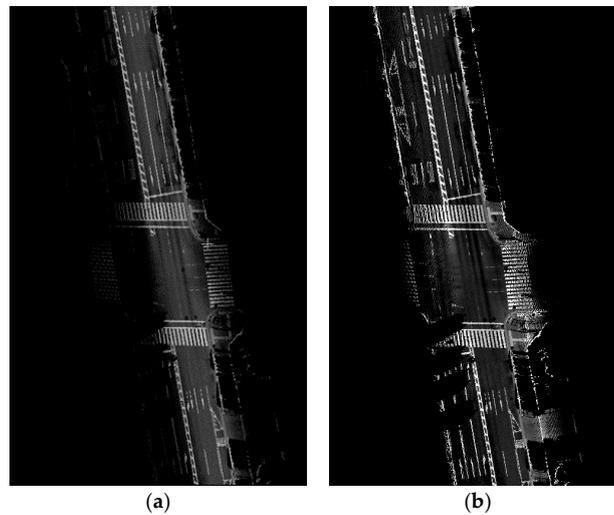


Figure 11. Intensity calibration of MMS point cloud: (a) original intensity of the MMS point cloud; and (b) calibrated intensity using the proposed method.

2.2.3. Road Marking Extraction

After calibrating the intensity, we should extract similar road markings from the MMS point cloud for the registration step. MMS road markings are derived by applying the adaptive thresholding procedure. Figure 12 shows a sample of the road markings extracted from an MMS point cloud, where the white points are the road markings. As can be seen, some curbs and vegetation close to the sidewalks are detected as road marking due to overestimation of the intensity in the previous step. However, those little misdetections do not affect the registration step, and can be neglected here.

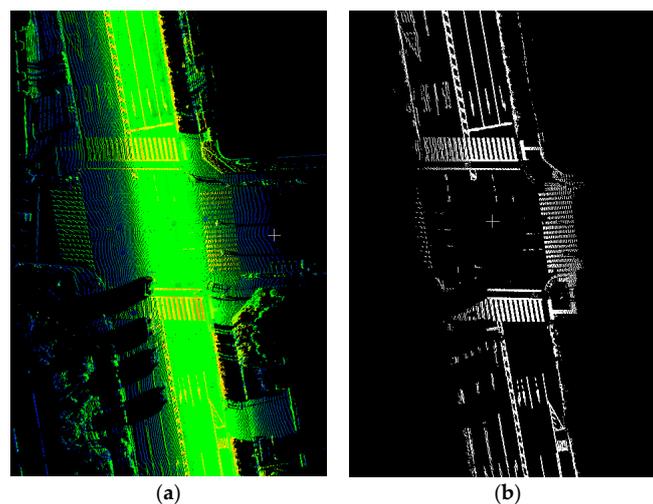


Figure 12. Road marking extraction from MMS point cloud: (a) original MMS point cloud (the color represents the original intensity); and (b) the result of road marking extraction.

2.3. 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 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.

2.3.1. 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 [52–54]. 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 (μ) 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 (Σ) is defined as

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

Figure 13a,b illustrates the representation of the estimated Gaussian distributions for 1D and 2D cells. Figure 13c shows a part of the aerial road markings, and the generated Gaussian mixture map is shown in Figure 13d. The probability of observing a road marking at a particular 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 represents the 2D coordinate (x, y) of the point.

The size of the NDT grid is an important parameter for the registration process. If the grid size selected is larger than the ideal dimension, the defined probability density function for the cell may not provide a good representation of the points included in the cell. In contrast, if the selected grid size is too small, each grid will capture only a few of the input road-marking points and only a few grids will have a Gaussian distribution [55]. Therefore, a small error will cause the input points to lie outside the corresponding reference grid.

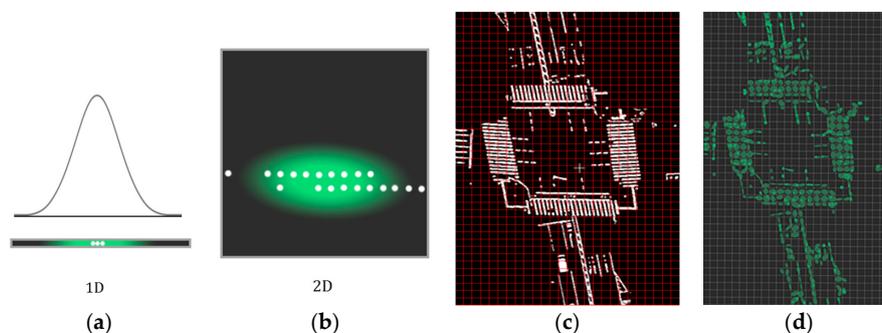


Figure 13. Generation of the Gaussian distribution from the reference data: (a) estimated 1D Gaussian distribution of the sample points; (b) estimated 2D Gaussian distribution of the sample points; (c) extracted aerial road markings; and (d) generated NDT map with a 2 m grid size.

2.3.2. 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 [56]. 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 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 surround 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 [57], which sometimes exceeds a few meters.

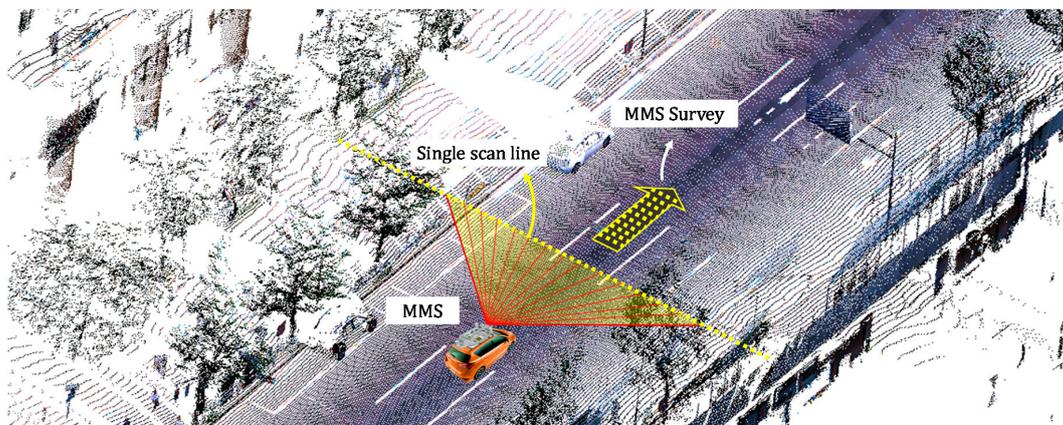


Figure 14. Illustration of single laser scan and MMS survey.

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 15.

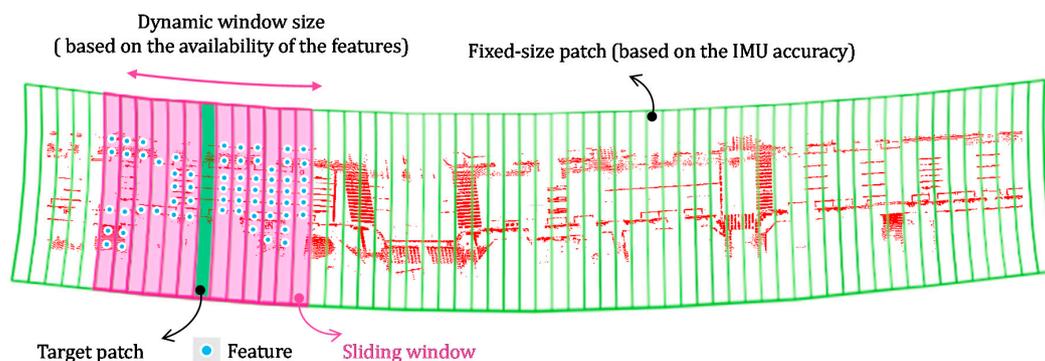


Figure 15. Concept of the dynamic sliding window.

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 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 paper, 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 15.

The window length is defined for each target patch based on three predefined factors: (1) patch length; (2) initial window length; and (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 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.

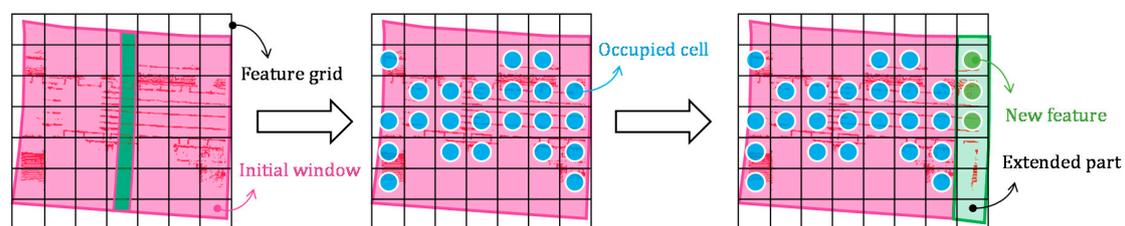


Figure 16. The procedure of defining the window length for the target patch. First, the initial window around the target patch is divided into a feature grid (**left**). Second, the occupied cells—those containing five or more points—are calculated (**middle**). The window is extended by adding new patches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

2.3.3. Registration

In this section, we describe the registration of the sliding window over the MMS point cloud (which is 3D) to the aerial road markings (which is 2D) using NDT matching at each step, which defines the transformation of the target patch in the original MMS survey. To obtain the rotation $R_z(\theta)$ and translation ($T_{x,y}$) of the 2D transformation matrix,

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

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 road markings.

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 [53]. 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 $\vec{x}_k \in W'_{t_i}$ 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:

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

Moreover, to optimize the cost function and achieve the best matching result iteratively, the Newton optimization method is employed [53].

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. Experiment and Analyses

3.1. MMS, Aerial Surveillance System, and Experimental Area

In this section, the details of the MMS and aerial surveillance system, the two primary data acquisition platforms used for the data collection, are described, and the experimental area is introduced.

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 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 precise vehicle self-localization. Two laser scanners, with 180° field of view, were configured to point up (pitch: +25°) and down (pitch: −25°). The laser scanners collect reflective luminance, which is a measure of the return strength of the laser pulse that generated the point. As we only need the road markings for the registration, any other setup that can densely scan the ground surface and capture the pavement markings is suitable for the experiments.

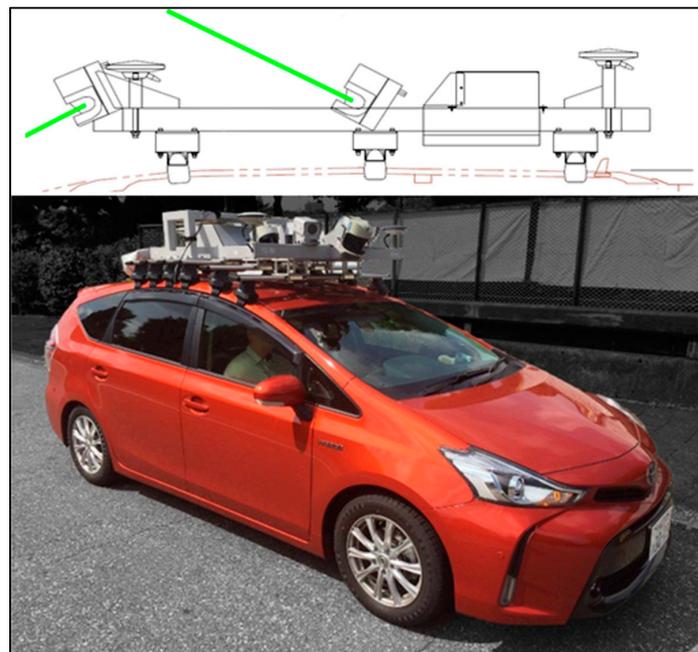


Figure 17. MMS system description: Mitsubishi Electric’s MMS-K320 (**bottom**); and the configuration of two SICK LMS-511 laser scanners and RTK GPS receivers (**top**).

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/m². Figure 18 shows the aerial system. Table 1 lists the details of the sensor platforms.



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

Table 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/s (1 unit)
		Range (max.)	65 m
	Camera	Viewing angle	180°
		No. mounted	3
		No. of pixels	5 megapixels
		Max capture rate	10 images/s
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 ²
	Camera	Field of view	$\sim 75^\circ$
		Manufacturer (Model)	Leica (RCD30)
		Mounting direction	Direct-down
		No. of pixels	80 MP (10,320 \times 7752 pixels)
		Maximum frame rate	0.8 fps
Localization platform	Forward overlap	60%	
	GSD * ⁷	12 cm/pixel	
	Manufacturer (Model)	Novatel (IMU-LN200)	
	Method	GNSS/IMU	

* 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 approximately 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.

3.1.3. Experimental Area

To evaluate the performance of the proposed system, the MMS and aerial data were acquired near Hitotsubashi, a dense urban area in the Chiyoda-ku area of central Tokyo, Japan (Figure 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 [58]. 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 12 June 2014 (a cloudy day in spring), and the MMS data were acquired on 20 April 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 12 June 2014 (the date of the aerial image acquisition). To simplify error interpretation, we 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.

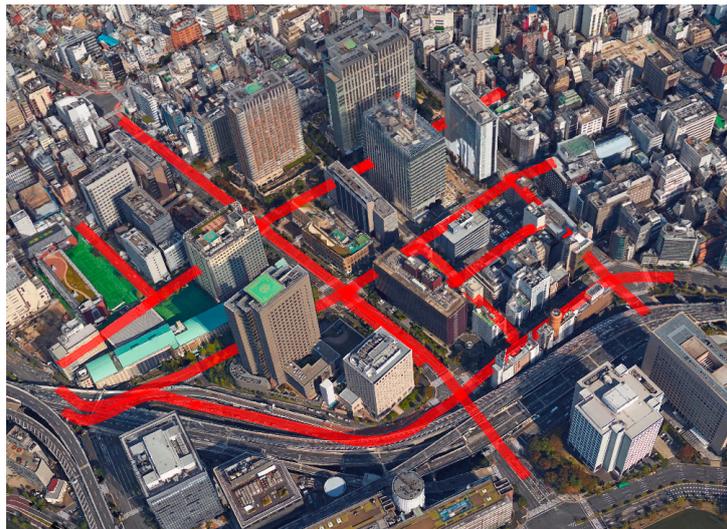


Figure 19. The experimental area around Hitotsubashi intersection, a dense urban area in Chiyoda-ku, Tokyo, Japan.

3.2. Experimental Parameters

This section provides an overview of the experimental parameters used for the final evaluation. Table 2 lists the parameters.

Table 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
NDT registration	Required feature count	400	Empirically defined
	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. Results

Nine MMS surveys were georeferenced using the proposed framework. The survey routes are shown in Figure 20. 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 s per meter per survey, which included the registration, visualization, georeferencing of the original data, and saving the updated data. We successfully calibrated all nine surveys using the proposed framework.

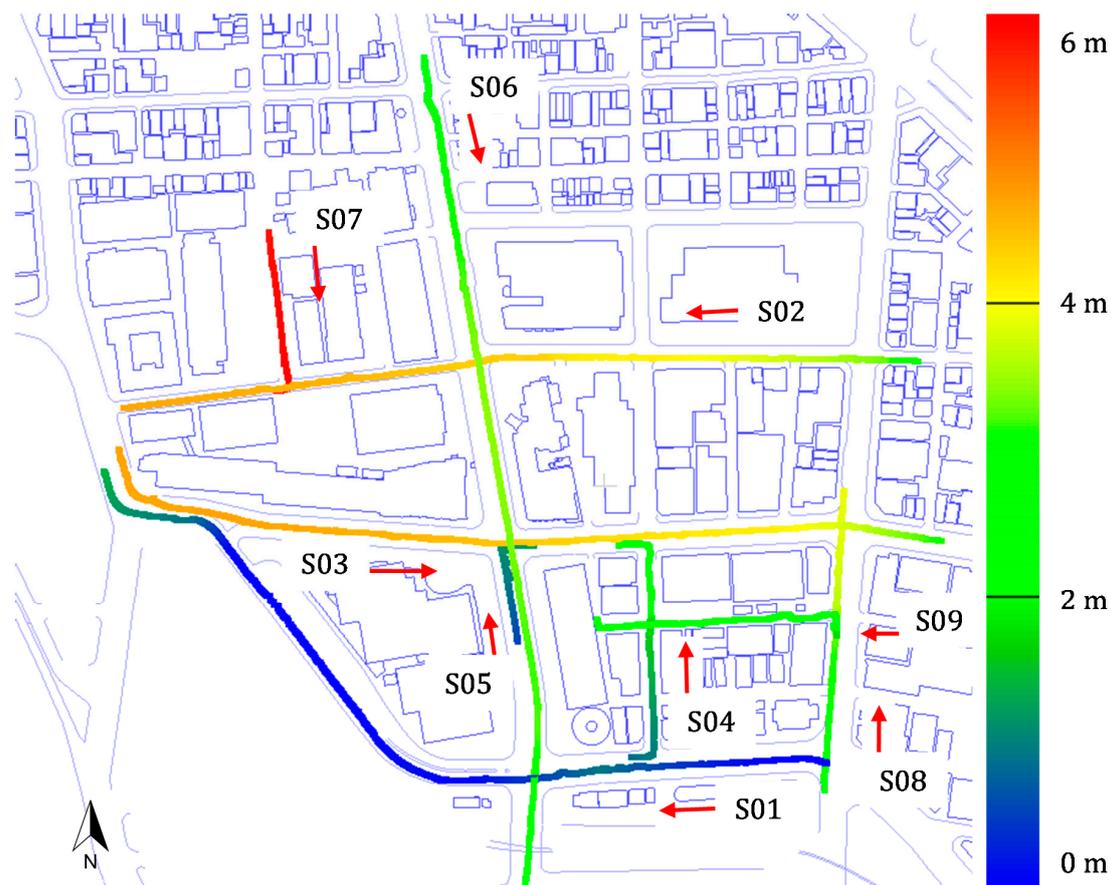


Figure 20. Routes of the nine surveys S01–S07 on the map. The colors show the estimated GPS error reported by MMS.

Figure 21 illustrates the performance of the proposed method for survey No. 6 visually. The survey's length was approximately 557 m. In this figure, the color of the MMS point cloud represents the intensity of the lidar reflection. In Figure 21b, the 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 21c).

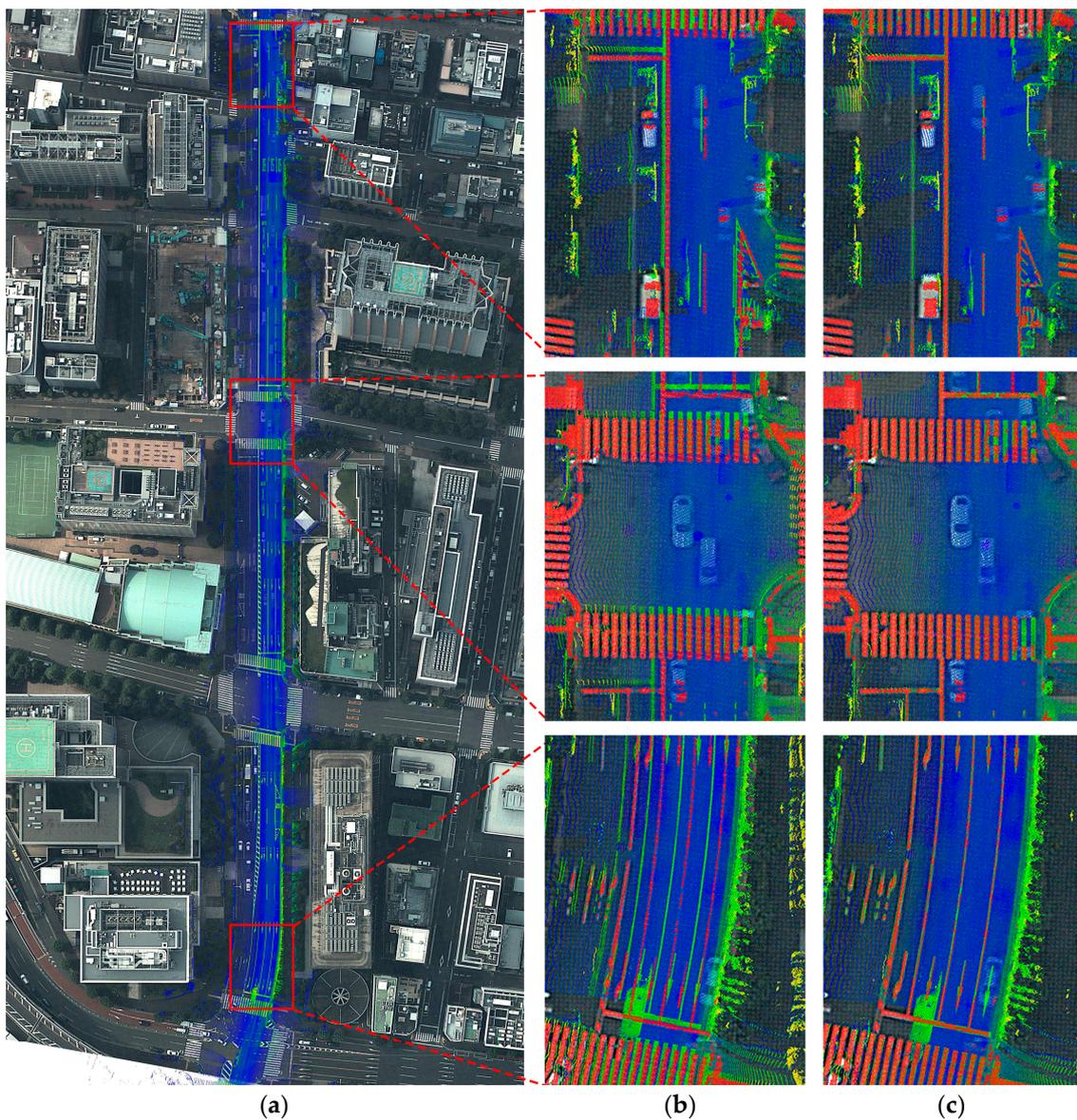


Figure 21. Visual evaluation of the proposed method for survey No. 6: (a) survey route of the MMS on the aerial image (after registration); (b) enlarged view of the original MMS point cloud before registration (the red points are aerial road markings); and (c) enlarged view of the MMS point cloud after applying the proposed method (the red points are aerial road markings).

4. 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 m, 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 m before the vehicle arrived back at the start point of the survey. Figure 22 shows the estimated GPS error for each of the nine surveys.

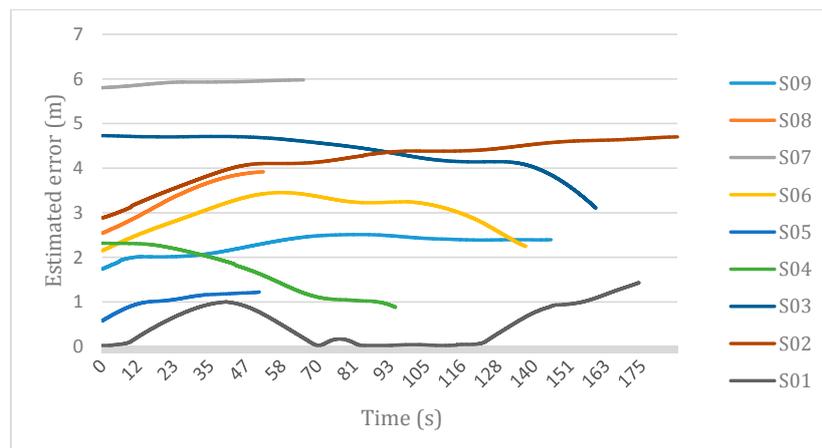


Figure 22. Estimated GPS error by MMS-K320 for each of the nine surveys.

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 23 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 georeferencing framework. For this purpose, the original GCPs were manually picked from the MMS point cloud using software provided by the Aisan Technology Co., Ltd. (Nagoya, Japan) to obtain the calibrated point cloud. The 2D error d was calculated for all VCPs using the following equation:

$$d_i = \sqrt{(X_{VCP_i} - x_i)^2 + (Y_{VCP_i} - y_i)^2}, \quad (13)$$

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.

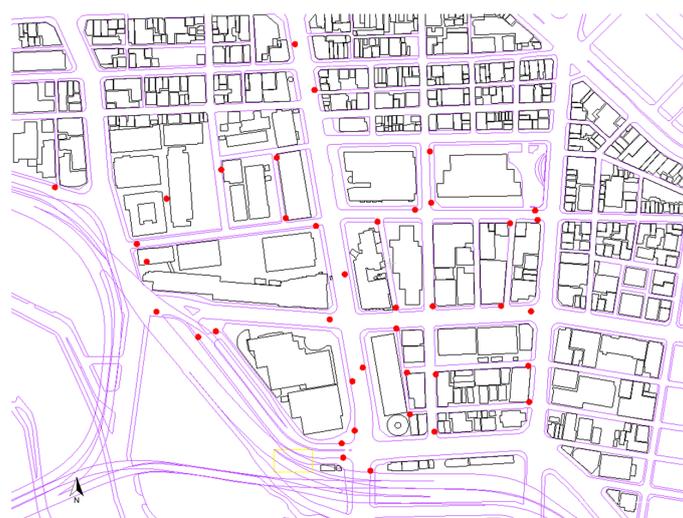


Figure 23. Distribution of the GCPs collected in the experimental area utilizing the total station survey. The GCPs are from the corner points of road signs, which are clearly captured by both the MMS camera and the point cloud.

Figure 24 shows a comparison of the proposed method, landmark updating and the original MMS data for survey No. 6, which extended for 557 m 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 summarizes the compared results.

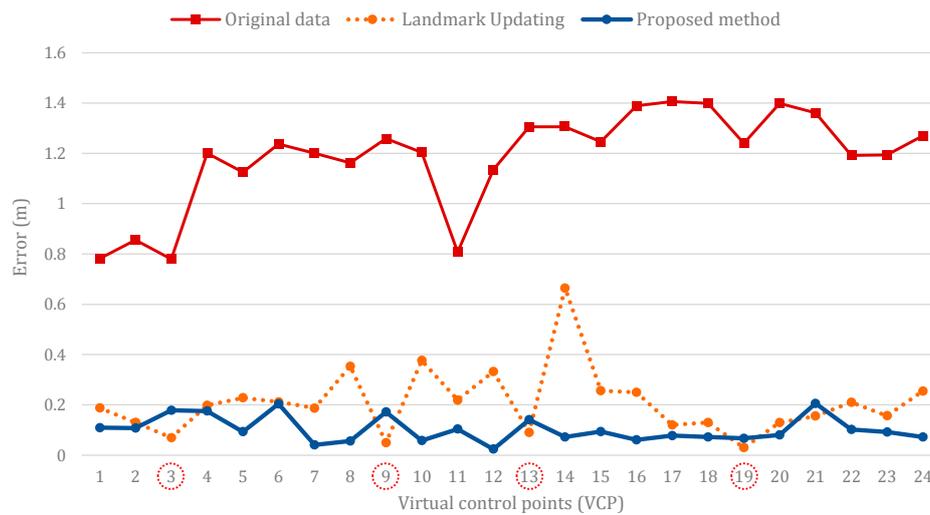


Figure 24. The 2D error of survey No. 6 in the original data, after the landmark update, and after the proposed method.

Table 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

In Figure 24, 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 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.

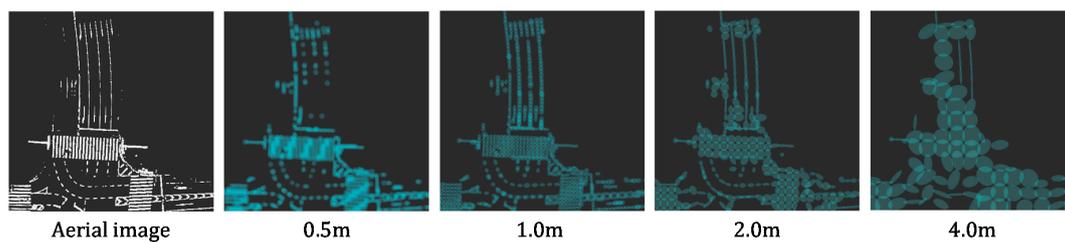
The proposed framework depends on several variables and parameters. In particular, the final registration step is sensitive to a variety of parameters such as the NDT grid size, the initial window length, the required feature count, and the patch length. We evaluated the sensitivity of the proposed framework 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 4 shows the registration result of survey No. 6 for various NDT grid sizes. As shown, grid size plays an important role in registration quality.

Table 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

* All parameters other than the NDT grid size were selected based on Table 2.

On the one hand, when the grid size is too small (e.g., 0.5 cm), some grids will 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 (i.e., road markings). Figure 25 shows the Gaussian mixture models generated using different grid sizes.

**Figure 25.** Generated Gaussian mixture models using different grid sizes.

The window length is another parameter that affects the registration performance. Figure 26 illustrates how the length of the dynamic sliding window changed for survey No. 6.

**Figure 26.** Length of the dynamic sliding window for survey No. 6

Tables 5 and 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) causes 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 would 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.

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

Item	Value			
	Proposed			
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

* All parameters other than the initial window length were chosen based on Table 2.

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, large feature counts (e.g., 600 or 800) will extend the window length disproportionately, which decreases the matching accuracy (Table 6).

Table 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 2.

Finally, the patch length is important to obtain a smooth point cloud from the registration. Figure 27 illustrates the georeferenced MMS point cloud using different patch lengths, where the initial window length was set to 30 m to evaluate only the patch length change effect on the registration. When the initial window length is defined by meters, the initial number of patches in a window decreases by increasing the patch size. As shown, long patches (e.g., 2 and 4 m) result in a non-smooth overlap between two consecutive patches (highlighted by red arrows). Although a smaller patch length is preferred, smaller patch lengths linearly increase the execution time. Accordingly, the patch length was set to 0.5 m, which provides both smooth registration and acceptable execution time.

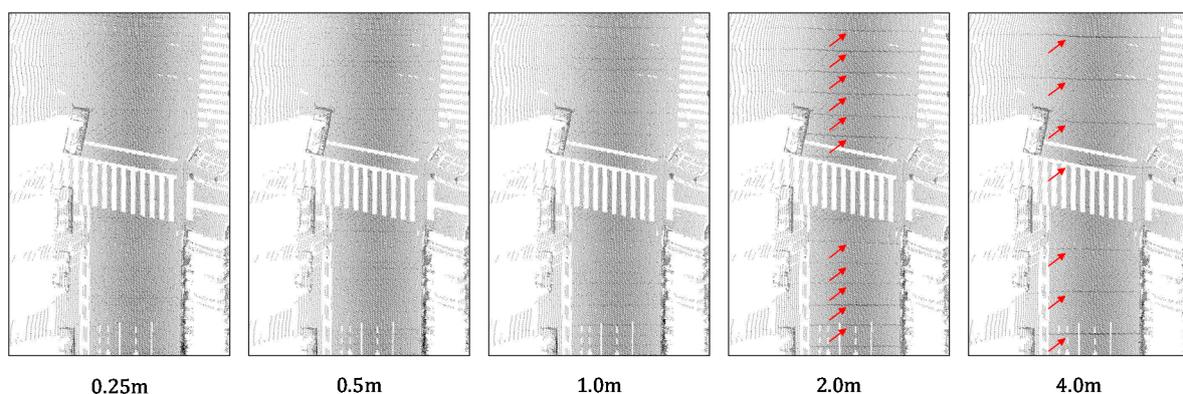


Figure 27. Effect of the patch length on the smoothness of the georeferenced MMS point cloud.

Figure 28 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 reported for the original data was 206 cm.

Figure 29 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.

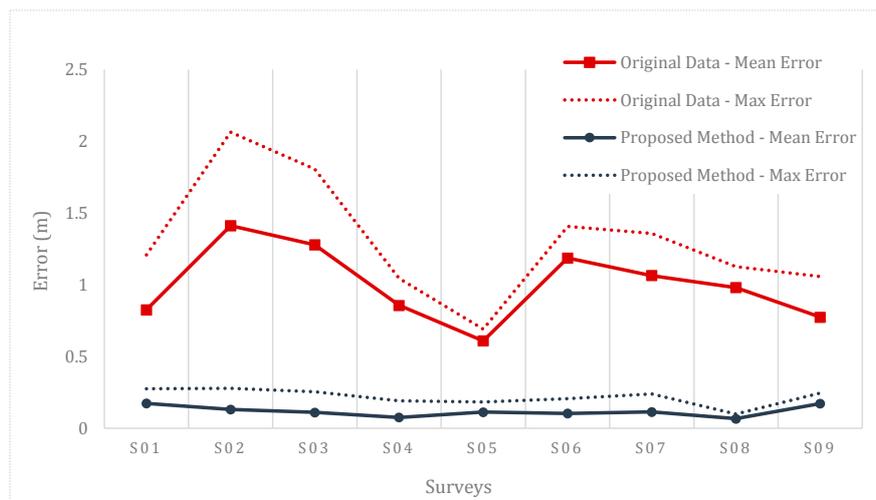


Figure 28. Total 2D error of the nine surveys in the original data and by the proposed method.

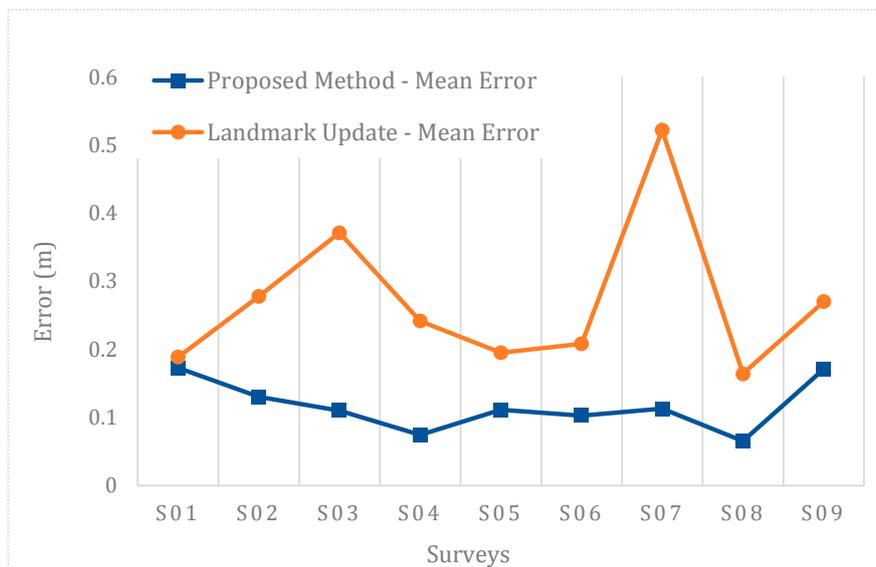


Figure 29. Total 2D error of the nine surveys by the landmark update method and by the proposed method.

Table 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. [35] is its accuracy. In contrast, although we could achieve a higher accuracy than Hussnain et al. [38], our method's main advantage is that the proposed framework is robust to temporary absence 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. We overcame this challenge by applying dynamic overlapping sliding windows for registration. The same group reported in Jende et al. [59] 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. However, the datasets used in these works are not the same, thus a fair comparison is not possible.

Table 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 Figure 13 of the cited paper. ** The reported numbers are extracted from Table 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.

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 30 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 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.



Figure 30. Areas where the road markings were repainted after 12 June 2014 (the date of the aerial image acquisition).

5. Conclusions

In this paper, 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 georeferencing 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 future work, we will consider a three-dimensional calibration of the MMS data based on the fusion of aerial images and ALS point clouds. After obtaining the 3D registration of the MMS data, it can be combined with the aerial data to create large-scale 3D urban models.

Author Contributions: Mahdi Javanmardi provided the core idea and conception of the proposed framework, designed the framework, analyzed the experiments, and wrote the manuscript. Ehsan Javanmardi contributed to the implementation of the MMS registration, helped revise the paper, and participated in all the discussions of the framework design and the experiment analysis. Yanlei Gu contributed to the calibration of the aerial image, discussions for the framework design, and helped revise the paper. Shunsuke Kamijo supervised this research project and participated in the discussions of the framework design and the analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rodríguez-Cuenca, B.; García-Cortés, S.; Ordóñez, C.; Alonso, M.C. Automatic Detection and Classification of Pole-Like Objects in Urban Point Cloud Data Using an Anomaly Detection Algorithm. *Remote Sens.* **2015**, *7*, 12680–12703. [[CrossRef](#)]
2. Gwon, G.P.; Hur, W.S.; Kim, S.W.; Seo, S.W. Generation of a Precise and Efficient Lane-Level Road Map for Intelligent Vehicle Systems. *IEEE Trans. Veh. Technol.* **2016**, *66*, 4517–4533. [[CrossRef](#)]
3. Kim, H.; Liu, B.; Goh, C.Y.; Lee, S.; Myung, H. Robust Vehicle Localization Using Entropy-Weighted Particle Filter-based Data Fusion of Vertical and Road Intensity Information for a Large Scale Urban Area. *IEEE Robot. Autom. Lett.* **2017**, *2*, 1518–1524. [[CrossRef](#)]
4. Hammoudi, K.; Dornaika, F.; Soheilian, B.; Vallet, B.; McDonald, J.; Paparoditis, N. Recovering Occlusion-Free Textured 3D Maps of Urban Facades by a Synergistic Use of Terrestrial Images, 3D Point Clouds and Area-Based Information. *Procedia Eng.* **2012**, *41*, 971–980. [[CrossRef](#)]
5. Cheng, L.; Wu, Y.; Tong, L.; Chen, Y.; Li, M. Hierarchical Registration Method for Airborne and Vehicle LiDAR Point Cloud. *Remote Sens.* **2015**, *7*, 13921–13944. [[CrossRef](#)]

6. Toschi, I.; Ramos, M.M.; Nocerino, E.; Menna, F.; Remondino, F.; Moe, K.; Poli, D.; Legat, K.; Fassi, F. Oblique Photogrammetry Supporting 3D Urban Reconstruction of Complex Scenarios. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Hannover, Germany, 6–9 June 2017; pp. 519–526.
7. Hammoudi, K.; McDonald, J. Enhancing Urban GIS Databases Through a Multi-view MMS Acquisition System for Street Feature Extraction. In Proceedings of the Annual GIS Ireland Conference, Irish Organization for Geographic Information (IRLOGI), Dublin, Ireland, 11 October 2012.
8. Papanoditis, N.; Papelard, J.; Cannelle, B.; Houzay, E. Stereopolis II: A Multi-Purpose and Multi-Sensor 3D Mobile Mapping System for Street Visualisation and 3D Metrology. *Revue Française Photogrammétrie Télédétection* **2012**, *200*, 69–79.
9. Sairam, N.; Nagarajan, S.; Ornitz, S. Development of Mobile Mapping System for 3D Road Asset Inventory. *Sensors* **2016**, *16*, 367. [[CrossRef](#)] [[PubMed](#)]
10. Zhu, L.; Hyypä, J. The Use of Airborne and Mobile Laser Scanning for Modeling Railway Environments in 3D. *Remote Sens.* **2014**, *6*, 3075–3100. [[CrossRef](#)]
11. Weinmann, M.; Weinmann, M.; Mallet, C.; Brédif, M. A Classification-Segmentation Framework for the Detection of Individual Trees in Dense MMS Point Cloud Data Acquired in Urban Areas. *Remote Sens.* **2017**, *9*, 277. [[CrossRef](#)]
12. Holopainen, M.; Kankare, V.; Vastaranta, M.; Liang, X.; Lin, Y.; Vaaja, M.; Yu, X.; Hyypä, J.; Hyypä, H.; Kaartinen, H.; et al. Tree Mapping Using Airborne, Terrestrial and Mobile Laser Scanning—A Case Study in a Heterogeneous Urban Forest. *Urban For. Urban Green.* **2013**, *12*, 546–553. [[CrossRef](#)]
13. Monnier, F.; Vallet, B.; Soheilian, B. Trees Detection from Laser Point Clouds Acquired in Dense Urban Areas by a Mobile Mapping System. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *1-3*, 245–250. [[CrossRef](#)]
14. Ranft, B.; Stiller, C. The Role of Machine Vision for Intelligent Vehicles. *IEEE Trans. Intell. Veh.* **2016**, *1*, 8–19. [[CrossRef](#)]
15. Wolcott, R.W.; Eustice, R.M. Robust LIDAR Localization Using Multiresolution Gaussian Mixture Maps for Autonomous Driving. *Int. J. Robot. Res.* **2017**, *36*, 292–319. [[CrossRef](#)]
16. Bengler, K.; Dietmayer, K.; Farber, B.; Maurer, M.; Stiller, C.; Winner, H. Three Decades of Driver Assistance Systems: Review and Future Perspectives. *IEEE Intell. Transp. Syst. Mag.* **2014**, *6*, 6–22. [[CrossRef](#)]
17. Jin, H.; Miska, M.; Chung, E.; Li, M.; Feng, Y. Road Feature Extraction from High Resolution Aerial Images Upon Rural Regions Based on Multi-Resolution Image Analysis and Gabor Filters. In *Remote Sensing—Advanced Techniques and Platforms*; InTech: Rijeka, Croatia, 2012.
18. Ellul, C.; Adjrard, M.; Groves, P. 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*; Copernicus GmbH: Göttingen, Germany, 2016; Volume IV-2/W1, pp. 171–178.
19. Yoneda, K.; Yang, C.; Mita, S.; Okuya, T.; Muto, K. Urban road localization by using multiple layer map matching and line segment matching. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 28 June–1 July 2015; pp. 525–530.
20. Choi, J. Hybrid Map-Based SLAM Using a Velodyne Laser Scanner. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 3082–3087.
21. Wolcott, R.W.; Eustice, R.M. Visual localization within LIDAR maps for automated urban driving. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 176–183.
22. Kaartinen, H.; Hyypä, J.; Kukko, A.; Jaakkola, A.; Hyypä, H. Benchmarking the Performance of Mobile Laser Scanning Systems Using a Permanent Test Field. *Sensors* **2012**, *12*, 12814–12835. [[CrossRef](#)]
23. Ishikawa, K.; Takiguchi, J.; Amano, Y.; Hashizume, T. A Mobile Mapping System for road data capture based on 3D road model. In Proceedings of the 2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control, Munich, Germany, 4–6 October 2006; pp. 638–643.
24. Lim, S.; Thatcher, C.A.; Brock, J.C.; Kimbrow, D.R.; Danielson, J.J.; Reynolds, B.J. Accuracy Assessment of a Mobile Terrestrial Lidar Survey at Padre Island National Seashore. *Int. J. Remote Sens.* **2013**, *34*, 6355–6366. [[CrossRef](#)]

25. Kajiwara, N.; Takiguti, J.; Hirokawa, R.; Shima, Y. Landmark Update of GPS/INS for Mobile Mapping System Using a Nearly Horizontal Single Camera and 3D Point-Cloud. In Proceedings of the International Symposium on GPS/GNSS, Tokyo, Japan, 11–14 November 2008.
26. González-Jorge, H.; Sánchez, J.M.; Díaz-Vilariño, L.; Puente, I.; Arias, P. Automatic Registration of Mobile LiDAR Data Using High-Reflectivity Traffic Signs. *J. Constr. Eng. Manag.* **2016**, *142*. [[CrossRef](#)]
27. Yao, L.; Wu, H.; Li, Y.; Meng, B.; Qian, J.; Liu, C.; Fan, H. Registration of Vehicle-Borne Point Clouds and Panoramic Images Based on Sensor Constellations. *Sensors* **2017**, *17*, 837. [[CrossRef](#)] [[PubMed](#)]
28. Iavarone, A.; Vagners, D. Sensor fusion: Generating 3D by combining airborne and tripod-mounted LIDAR data. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Tarasp, Switzerland, 24–28 February 2003; Volume XXXIV-5/W10.
29. Cheng, L.; Tong, L.; Wu, Y.; Chen, Y.; Li, M. Shiftable Leading Point Method for High Accuracy Registration of Airborne and Terrestrial LiDAR Data. *Remote Sens.* **2015**, *7*, 1915–1936. [[CrossRef](#)]
30. Yang, B.; Zang, Y.; Dong, Z.; Huang, R. An automated method to register airborne and terrestrial laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* **2015**, *109*, 62–76. [[CrossRef](#)]
31. Kedzierski, M.; Fryskowska, A. Terrestrial and Aerial Laser Scanning Data Integration Using Wavelet Analysis for the Purpose of 3D Building Modeling. *Sensors* **2014**, *14*, 12070–12092. [[CrossRef](#)] [[PubMed](#)]
32. Teo, T.-A.; Huang, S.-H. Surface-Based Registration of Airborne and Terrestrial Mobile LiDAR Point Clouds. *Remote Sens.* **2014**, *6*, 12686–12707. [[CrossRef](#)]
33. Cheng, L.; Tong, L.; Li, M.; Liu, Y. Semi-Automatic Registration of Airborne and Terrestrial Laser Scanning Data Using Building Corner Matching with Boundaries as Reliability Check. *Remote Sens.* **2013**, *5*, 6260–6283. [[CrossRef](#)]
34. Polewski, P.; Erickson, A.; Yao, W.; Coops, N.; Krzystek, P.; Stilla, U. Object-Based Coregistration of Terrestrial Photogrammetric and ALS Point Clouds in Forested Areas. In *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*; Copernicus GmbH: Göttingen, Germany, 2016; Volume III-3, pp. 347–354.
35. Kümmerle, R.; Steder, B.; Dornhege, C.; Kleiner, A.; Grisetti, G.; Burgard, W. Large scale graph-based SLAM using aerial images as prior information. *Auton. Robots* **2011**, *30*, 25–39. [[CrossRef](#)]
36. Gruen, A.; Huang, X.; Qin, R.; Du, T.; Fang, W.; Boavida, J. Adriano Oliveira Joint processing of UAV imagery and terrestrial MMS data for very high resolution 3D city modeling. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-1/W2*. [[CrossRef](#)]
37. Tournaire, O.; Soheilian, B.; Paparoditis, N. Towards a Sub-decimeter Georeferencing of Ground-based Mobile Mapping Systems in Urban Areas: Matching Ground-based and Aerial-Based Imagery Using Roadmarks—Semantic Scholar. In Proceedings of the ISPRS Commission I Symposium, Marne-la-Vallée, France, 4–6 July 2006.
38. Hussnain, Z.; Oude Elberink, S.; Vosselman, G. Automatic Feature Detection, Description and Matching from Mobile Laser Scanning Data and Aerial Imagery. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B1*, 609–616. [[CrossRef](#)]
39. Jende, P.; Peter, M.; Gerke, M.; Vosselman, G. Advanced Tie Feature Matching for the Registration of Mobile Mapping Imaging Data and Aerial Imagery. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B1*, 617–623. [[CrossRef](#)]
40. Jende, P.; Hussnain, Z.; Peter, M.; Oude Elberink, S.; Gerke, M.; Vosselman, G. 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*; Copernicus GmbH: Göttingen, Germany, 2016; Volume XL-3-W4, pp. 19–26.
41. Javanmardi, M.; Javanmardi, E.; Gu, Y.; Kamijo, S. A Novel Approach for Post-Calibration of Mobile Mapping Systems Using Intensity Reflection and Airborne Imagery. In Proceedings of the Transportation Research Board 96th Annual Meeting, Washington, DC, USA, 8–12 January 2017.
42. Li, Y.; Li, L.; Li, D.; Yang, F.; Liu, Y. A Density-Based Clustering Method for Urban Scene Mobile Laser Scanning Data Segmentation. *Remote Sens.* **2017**, *9*, 331. [[CrossRef](#)]
43. Meng, X.; Currit, N.; Zhao, K. Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. *Remote Sens.* **2010**, *2*, 833–860. [[CrossRef](#)]
44. Yang, B.; Huang, R.; Dong, Z.; Zang, Y.; Li, J. Two-Step Adaptive Extraction Method for Ground Points and Breaklines from Lidar Point Clouds. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 373–389. [[CrossRef](#)]

45. Guan, H.; Li, J.; Yu, Y.; Ji, Z.; Wang, C. Using Mobile LiDAR Data for Rapidly Updating Road Markings. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2457–2466. [[CrossRef](#)]
46. Zhang, W.; Qi, J.; Wan, P.; Wang, H.; Xie, D.; Wang, X.; Yan, G. An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. *Remote Sens.* **2016**, *8*, 501. [[CrossRef](#)]
47. Kaasalainen, S.; Jaakkola, A.; Kaasalainen, M.; Krooks, A.; Kukko, A. Analysis of Incidence Angle and Distance Effects on Terrestrial Laser Scanner Intensity: Search for Correction Methods. *Remote Sens.* **2011**, *3*, 2207–2221. [[CrossRef](#)]
48. Kukko, A.; Kaasalainen, S.; Litkey, P. Effect of Incidence Angle on Laser Scanner Intensity and Surface Data. *Appl. Opt.* **2008**, *47*, 986–992. [[CrossRef](#)] [[PubMed](#)]
49. Kashani, A.G.; Olsen, M.J.; Parrish, C.E.; Wilson, N. A Review of LIDAR Radiometric Processing: From Ad Hoc Intensity Correction to Rigorous Radiometric Calibration. *Sensors* **2015**, *15*, 28099–28128. [[CrossRef](#)] [[PubMed](#)]
50. Jutzi, B.; Gross, H. Normalization of Lidar Intensity Data Based on Range and Surface Incidence Angle. In Proceedings of the Laserscanning 2009, Paris, France, 1–2 September 2009; pp. 213–218.
51. Soudarissanane, S.; Lindenbergh, R.; Menenti, M.; Teunissen, P. Incidence Angle Influence on the Quality of Terrestrial Laser Scanning Points. In Proceedings of the Laserscanning 2009, Paris, France, 1–2 September 2009; pp. 83–88.
52. Magnusson, M.; Vaskevicius, N.; Stoyanov, T.; Pathak, K.; Birk, A. Beyond Points: Evaluating Recent 3D Scan-Matching Algorithms. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 3631–3637.
53. Magnusson, M. The Three-Dimensional Normal-Distributions Transform—An Efficient Representation for Registration, Surface Analysis, and Loop Detection. Ph.D. Thesis, Örebro University, Örebro, Sweden, 2009.
54. Biber, P.; Strasser, W. The Normal Distributions Transform: A New Approach to Laser Scan Matching. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), Las Vegas, NV, USA, 27–31 October 2003; Volume 3, pp. 2743–2748.
55. Kaminade, T.; Takubo, T.; Mae, Y.; Arai, T. The Generation of Environmental Map Based on a NDT Grid Mapping—Proposal of Convergence Calculation Corresponding to High Resolution Grid. In Proceedings of the 2008 IEEE International Conference on Robotics and Automation, St. Louis, MO, USA, 10–15 October 2008; pp. 1874–1879.
56. Yoshida, Mi.; Yoshida, Ma. The Latest Trends in Mobile Mapping System. *Mitsubishi Electr. Adv. Mag.* **2014**, *148*, 2–5.
57. Haala, N.; Peter, M.; Cefalu, A.; Kremer, J. Mobile Lidar Mapping for Urban Data Capture. In Proceedings of the 14th International Conference Virtual System Multimedia, Limassol, Cyprus, 20–26 October 2008; pp. 95–100.
58. Miura, S.; Hsu, L.-T.; Chen, F.; Kamijo, S. GPS Error Correction with Pseudorange Evaluation Using Three-Dimensional Maps. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 3104–3115. [[CrossRef](#)]
59. Jende, P.; Nex, F.; Gerke, M.; Vosselman, G. 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*; Copernicus GmbH: Göttingen, Germany, 2017; Volume XLII-1-W1, pp. 317–323.

