2.5D Layered Sub-Image LIDAR Maps for Autonomous Driving in Multilevel Environments

Mohammad Aldibaja *, Naoki Suganuma and Ryo Yanase

The Advanced Mobility Research Institute, Kanazawa University, Kanazawa 920-1192, Japan
* Correspondence: amroaldibaja@staff.kanazawa-u.ac.jp; Tel.: +81-76-234-4728

Abstract: This paper proposes a reliable framework to map multilevel road structures in the 2D image domain called layered sub-image maps (LSM). The road is divided into a set of sub-areas providing IDs in the XY plane. Each sub-area is decomposed into several layered images using LIDAR intensity and elevation data to form a 2.5D map image. The layered elevation images are given IDs in the Z plane to represent the height of the contained road features in meter-order whereas the elevation pixels indicate the cm-order of the road slope in the range of 200 cm. The layered intensity images are then created to describe the road surface in conjunction with the number of the layered elevation images and the corresponding pixel distributions. A significant map retrieval strategy during autonomous driving has been designed based on the LSM implementation tactic and the IDs in the XYZ plane. The system’s reliability has been proved by a unique localization module to localize an autonomous vehicle in a challenging multilevel environment consisting of four stacked loops with an average accuracy of 5 cm in lateral, longitudinal and altitudinal directions.

Keywords: LIDAR; multilevel environments; 2.5D intensity-elevation maps; autonomous vehicles; multilayer roads

1. Introduction

Mapping in multilevel environments is a challenging and important requirement to deploy autonomous vehicles in modern cities that are packed with longitudinal bridges and underground tunnels. The mapping tactic in such environments including map creation, places identification in the absolute coordinate system (ACS) and storing strategy plays the main role in designing the localization module. This is because the maps should be completely retrieved for each layer based on the vehicle position in the XYZ plane during autonomous driving. The matching process between the map and sensory data can be then conducted safely to correct the localization errors [1–4]. Therefore, mapping and localization in multilayer road structures are considered as a challenging demand in levels four and five of autonomous driving and are rarely discussed in the autonomous vehicle research field.

Triebel has proposed a method called a multi-level surface map (MLS) to classify environment patches into traversable, non-traversable and vertical [5]. The idea is to divide environments into cells that can encode many surfaces based on the mean and the variance of the height as well as the depth of the vertical extension [6]. This facilitates distinguishing the bridge and underpass patches in a local LIDAR point cloud. In addition, it significantly enhances the matching result by iterative closest point (ICP) to correct the relative-positions errors of scanning the same places and creates maps using SLAM technologies [7,8]. The generated maps have been used for enabling path planning and localization in [9] with reliable results of surfing in a ground parking lot. Kummerle has used MLS maps to scan a multilayer parking lot and conduct autonomous localization accordingly [10]. However, storing depth, height and variance of cells (20 cm) in each point cloud is a high-cost process and occupies a large size of memory to create large maps. In
addition, it increases the processing time to retrieve maps during autonomous driving because of the need to add an ID to each cell. Furthermore, MLS has been designed to operate in the 3D point cloud domain to provide the shape and distribution patterns of the classified features. Consequently, the MLS maps are very applicable in small areas because the size of cells should be preserved small to ensure sufficient resolution of the environment representation. Accordingly, we extend our previous work [11] and propose a reliable strategy to map multilevel environments in the 2D image domain and represent the intensity and elevation of road surfaces, i.e., layered sub-image maps (LSM). In addition, a 3D localization module is explained to significantly utilize LSMs to estimate the \( xyz \) vehicle position during autonomous driving. A complex highway multilayer road junction in Tokyo was used to emphasize the robustness, scalability and efficiency of the proposed system to safely retrieve maps in layers and localize vehicles accordingly.

2. Previous Work

In the previous work [11], we proposed a 2D mapping strategy of road surfaces while considering the point cloud accumulation, the storing size of map images and the identification of the images in an ACS. To address these issues, the dynamic objects in the LIDAR point clouds are filtered out based on the velocity (more than 0.4 km/h) and a height threshold. The height threshold is set to 0.3 m to preserve the points of stationary environmental features such as painted landmarks and curbs as well as the lower parts of trees, road edges, barriers, poles, traffic lights and so on. A 2D LIDAR intensity frame is then generated by converting the filtered point cloud into an image based on the laser beam reflectivity as shown in Figure 1a. This is achieved using a series of transformations between an ACS, a vehicle, LIDAR and image coordinate systems. The LIDAR frames are then accumulated in an intensity image based on the vehicle trajectory as illustrated in Figure 1b–d. The accumulation process is terminated by exceeding the number of pixels of a fixed threshold \( \lambda \) as demonstrated in Figure 1e and a node is produced accordingly. The top-left corner \( C_{TF} \) is used to identify the node in the ACS and the coordinates are determined by the maximum \( x_y \) vehicle positions inside the driving area as indicated in Figure 1e. The node’s intensity image is then divided into a set of sub-images with fixed width and height \( s_w \times s_h \). The top-left corner is utilized to label the sub-images by global IDs \( (x, y) \) as in Equation (1) and Figure 1f. The IDs are used to save the sub-images into a hard disc in a BMP image compression format. These technical steps are continuously applied to create new nodes and extend the map size in the image domain.

\[
ID(x) = \text{floor}(C_{TF}(x)/s_w) \\
ID(y) = \text{ceil}(C_{TF}(y)/s_h)
\]  

(1)
3. Problem Statement

The mapping system in [11] has been used to generate maps for more than 7000 km in different cities in Japan. A GNSS/INS-RTK (GIR) system has been used to estimate the vehicle position during the data collection phase and calculate the node top-left corners’ $xy$ coordinates in the ACS. Accordingly, the intensity maps are generated and stored by sub-images with the corresponding XY-IDs. A robust localization system has been proposed to localize the vehicle using the intensity maps in the XY plane [12]. The need to perform autonomous driving in multilayer road structures has then emerged and the problem of encoding different layers in the same 2D sub-image has been observed [13,14]. Accordingly, considerable localization errors may occur in the overlapped zones such as bridges, tunnels.
and multi-level parking lots because they share the same IDs \((x, y)\). Figure 2a shows an environment of three road layers, i.e., a ground tunnel, a bridge tunnel and a mountain road. The overlapping of these layers in the 2D sub-image completely distorts the road structure representation as illustrated in Figure 2b. Consequently, the localization module fails to estimate the vehicle position because the sensory data (observation image) encode only the road surface of a unique layer during autonomous driving. Figure 2c–e shows that the map image is overlapped by the observation image in each layer based on the matching score. The red and yellow circles imply the actual and estimated vehicle positions, respectively. The red rectangle demonstrates the correct overlapping of the observation image based on the actual position. A lateral deviation occurs in the ground-tunnel layer, i.e., the middle-dashed lines do not match those in the map image. A longitudinal deviation is produced in the bridge-tunnel layer that indicates the road context is not aligned properly. Finally, a perfect match is obtained in the mountain layer because of the unique road pattern compared to the parallel patterns of the other two layers. Therefore, the road layers must automatically be distinguished and registered in the maps and the localization system must be modified to estimate the vehicle position in the XYZ plane, accordingly.

Figure 2. (a) Multilevel environment consisting of three layers: ground-tunnel, bridge-tunnel and mountain surface. (b) The three layers are accumulated in the same map image [8]. (c–e) Map image in (b) overlapped by LIDAR image based on the matching score in the three layers, respectively. Red circle and rectangle show the accurate actual location whereas yellow circle indicates the estimated position.
4. Proposed Solution

4.1. Mapping Multilayer road Structures

The above problem can be solved by separating the road layers in the map images based on the altitudinal values in the ACS. We modified the mapping system to generate maps in two representations: intensity and elevation, i.e., 2.5D maps [15–17]. The intensity image describes the road surface in grayscale pixels whereas the elevation image encodes the height value of each pixel in the real world. Therefore, the same technical steps in Section 2 are simultaneously applied to produce the elevation images of the nodes in a float format as particularly detailed below.

Given a rotation matrix \( R \) by the IMU system at a vehicle position \( p \), the index \((u, v)\) of a pixel can be calculated using Equations (2)–(4) with respect to the pixel resolution \( \sigma \).

\[
p' = \begin{bmatrix} p'_x \\ p'_y \\ p'_z \end{bmatrix} = RP \\
\]

\[
u = \frac{S_w}{2} + \frac{p'_x}{\sigma} \quad \text{(3)}
\]

\[
v = \frac{S_h}{2} + \frac{p'_y}{\sigma} \quad \text{(4)}
\]

The corresponding height of \( p' \) in the is obtained by adding the vehicle \( z \) position as in Equation (5).

\[
z_{u,v} = p'_z + z_v \quad \text{(5)}
\]

Thus, the LIDAR elevation frame is formed at each vehicle position and accumulated in an elevation map image as demonstrated for a bridge-underpass environment in Figure 3a. The environment is encoded by three nodes based on \( \lambda \) and merged in the ACS using the top-left corner’s coordinates in Figure 3b. The common area between the bridge and underpass occurs in the first and third nodes as demonstrated in Figure 3c. The underpass is first saved in the sub-images of the first node with the IDs \((x, y)\). Based on the vehicle trajectory in Figure 3a, the bridge segment in the third node is used to update the saved sub-images of the first nodes because they share the same IDs \((x, y)\), i.e., loop closure events [18]. Accordingly, the two layers are encoded in the same sub-image and distort the road representation in the XY and Z planes as illustrated in Figure 3d. In order to solve this problem, we considered the identification of sub-images in the Z plane by providing IDs \( z \) and mapping the road layers individually.

Storing elevation maps is very tricky and critical because of the need to include float values. In addition, an elevation map image may contain a particular set of pixels that demonstrate a considerable change in the height values because of the road slope. Therefore, identifying the road layers by IDs \( z \) must be addressed properly.

The ACS is virtually sampled into layers according to an altitudinal threshold \( \beta \). The threshold is selected to be less than the standard norm distance between bridge and ground road surfaces, e.g., 2 m. Based on \( \beta \), each layer can be given an ID \( z \) to represent the altitudinal range in the meter order. The height of the environmental features in each layer is then encoded by an elevation image in the cm-order with respect to the given ID \( z \) and the road slope. Finally, the pixels of the road surface are integrated and distributed in a set of layered intensity images according to the feature existence in the layered elevation images.

The simplest strategy to achieve the above tactic is to apply the above steps to each sub-image and create the LSM. The sub-image of the common area in the first node is initially considered. The maximum and minimum altitudinal values \( Z_{\text{max}} \) and \( Z_{\text{min}} \) are calculated and the corresponding two boundary layers in ACS are determined using Equation (6).
\begin{align*}
ID(z)_{\text{max}} &= \text{floor}(z_{\text{max}}/\beta) \\
ID(z)_{\text{min}} &= \text{floor}(z_{\text{min}}/\beta)
\end{align*}

Based on the two boundary layers and \( \beta \), the in-between layers are simply initialized with the corresponding IDs \( z \) as demonstrated in Figure 4a in the pink polygon. A layer’s plane \( i \) can mathematically be represented by Equation (7) according to the normal vector \([n_x, n_y, n_z]^T = [0, 0, 1]^T\).

\[ n_x(x - x_i^n) + n_y(y - y_i^n) + n_z(z - ID(z)_i) = 0 \]

where \([x, y, z]^T\) is an arbitrary position in the plane with the altitudinal range of IDs \([z_i] + \beta\). Hence, the pixels in the layered elevation sub-image are classified and distributed on the created planes and the coordinates of a feature in an \( i \)th layer are restored in the ACS using Equation (8).

\[
\begin{align*}
x &= (S_x \ast \sigma \ast ID(x)) + \sigma \ast u \\
y &= (S_y \ast \sigma \ast ID(y)) + \sigma \ast v \\
z &= (255 \ast \gamma \ast ID(z)) + \gamma \ast w
\end{align*}
\]

The corresponding intensity sub-image of the road surface is decomposed into the same number of layers as demonstrated in Figure 4b. The road surface pixels are then assigned according to the distribution in each elevation layer. The layered elevation and intensity images are saved with the IDs \([x, y, z]\). These IDs are effectively utilized to update the saved images and automatically add new layers to encode different levels of the same road structures, e.g., the bridge segment is added after creating the third node and represented by three planes as surrounded by the cyan polygon.

**Figure 3.** (a) Ground and bridge road structure represented by an elevation map showing the vehicle trajectory. (b) Intensity map consisting of three nodes. (c) Common area between bridge and ground road surfaces. (d) Intensity (top) and elevation (bottom) images of overlapped area.
created planes and the coordinates of a feature in an $i^{th}$ layer are restored in the ACS using Equation (8).

$$
\begin{align*}
    x &= (S_{i4} \ast \sigma \ast I_{D}(x)) + \sigma \ast u \\
    y &= (S_{i5} \ast \sigma \ast I_{D}(y)) + \sigma \ast v \\
    z &= (255 \ast \gamma \ast I_{D}(z)) + \gamma \ast w \\
\end{align*}
$$ (8)

Figure 4. Ground and bridge road structure shown in Figure 3 are mapped according to LSM. (a) The elevation image of the ground road surface is first encoded and divided into two layers in the pink polygon. The map is then extended in the Z direction based on the Z-IDs to contain the bridge layer images. (b) The corresponding intensity images are created according to the pixel distribution in the layered elevation images in (a).

4.2. Multilayer Map Retrieval during Autonomous Driving

Retrieving map during autonomous driving is a difficult issue because it needs to be strategized according to the localization system. Given a vehicle pose, the ID $(x, y)$ of the driving area in the ACS can be obtained by Equation (1) as illustrated in Figure 5a, i.e., $CTF$ is replaced with the vehicle pose’s $xy$ coordinates. The three neighboring areas are identified by frequently adding one unit to the driving area ID $(x, y)$. The four areas create a virtual plane that is rotated based on the pitch and roll angles $\text{Rot}$ and elevated by the $z$ vehicle position. The plane is represented by the normal vector $[n_x, n_y, n_z]^T = \text{Rot}[0, 0, 1]^T$. 

![Diagram](image_url)
Another two edge planes (lower and upper) passing through the points of \((x_i^0 + \frac{n_x}{2}n_x, y_i^0 + \frac{n_y}{2}n_y, z_i^0 + \frac{n_z}{2}n_z)\) and \((x_i^0 + \frac{n_x}{2}n_x, y_i^0 + \frac{n_y}{2}n_y, z_i^0 - \frac{n_z}{2}n_z)\) are assumed to determine the altitudinal range of each pixel in the plane. Accordingly, the layered elevation sub-images in the four areas are restored based on the IDs \((z)\) in the altitudinal range. The pixels in the virtual plane are then assigned by averaging the elevation values between the edge planes as shown in Figure 5b. The resultant plane is called the large elevation map image. This image is cut off around the vehicle with a fixed size to create the elevation map image. This image is cut off to localize the vehicle in the Z direction. Based on the pixel distribution in the elevation map image and IDs \((z)\), the corresponding pixels in the layered intensity sub-images are averaged to form the intensity map image and express the road surface as demonstrated in Figure 5c. Likewise, the image is cut off to obtain the intensity map image that will be used to localize the vehicle in the XY plane.

![Image](https://example.com/image.png)

Figure 5. Map retrieval strategy. (a) Restoring four sub-images of the ground layer in Figure 3 based on the vehicle pose and creating lower and upper virtual planes. (b) Elevation map image by merging the lower two layered elevation images in Figure 4a. (c) Intensity map image of the ground road surface.

5. LSM Based Localization in Multilevel Environments

The localization system is a critical pillar to enable autonomous driving. The vehicle must be localized in the longitudinal direction with respect to the road context, in the lateral direction with respect to the lane lines and in the altitudinal direction with respect to the road structure [19–21]. As safety is a prime priority, dead reckoning (DR) is employed to estimate the vehicle position \(X_{i,\text{DR}}\) using on-board sensors [22,23]. This is achieved by integrating the time interval \(\Delta t\) and the vehicle velocity \(v\) to the previous position \(X_{i-1,\text{DR}}\) as clarified in Equation (9).

\[
X_{i,\text{DR}} = X_{i-1,\text{DR}} + v_i \Delta t
\]

(9)

The DR accuracy is inversely proportional to the distance from the initial position. Therefore, the accuracy is recovered to precisely estimate the vehicle position \(X_{i,\text{ACS}}\) by a translation offset \(\Delta X_{i,\text{ACS,DR}}\) in Equation (10) based on matching the elevation-intensity map images and the online LIDAR point clouds as detailed in the next section.

\[
X_{i,\text{ACS}} = X_{i,\text{ACS,DR}} + \Delta X_{i,\text{ACS,DR}}
\]

(10)
5.1. Localization in XY Plane

The DR estimation is used to retrieve the LSM as explained in Section 4.2. The LIDAR encodes the surrounding environment according to the actual position of the vehicle in the real world. The LIDAR point clouds are down-sampled and converted into an intensity image to describe the road surface on the grayscale level. The intensity image is utilized to calculate the similarity score \( R(\Delta x, \Delta y) \) with the intensity map image using zero normalized cross correlation (ZNCC) as mathematically explained in (11) and demonstrated in Figure 6.

\[
R(\Delta x, \Delta y) = \frac{\sum_{i=0}^{N-1} (z_i - \bar{z}) (m_{i}(\Delta x, \Delta y) - \bar{m})}{\sqrt{\sum_{i=0}^{N-1} (z_i - \bar{z})^2} \sqrt{\sum_{i=0}^{N-1} (m_i(\Delta x, \Delta y) - \bar{m})^2}}
\]  

(11)

Figure 6. LSM based localization system in XY plane. Lateral and longitudinal offsets are first calculated in XY plane using image matching between intensity map and LIDAR images. This equation provides a matching matrix to indicate the location of the LIDAR image in the map image by the highest peak. As localization is not a deterministic process, the estimations must be tracked in a probabilistic framework. Therefore, the matching matrix...
is multiplied by the prior probability $P_t(\Delta x, \Delta y)$ to obtain the posterior probability $P_t$ as clarified in Equation (12).

$$P_t(\Delta x, \Delta y) = \eta(R_t(\Delta x, \Delta y) + 1)P_t(\Delta x, \Delta y)$$

(12)

The offsets $X_{L,ACS,DR}$ are then calculated based on the posterior probability and the corrected vehicle position is obtained accordingly, using Equation (10).

### 5.2. Localization in Z Plane

Based on $\Delta X_{L,ACS,DR}$ the retrieved MLS is cut off to be in the same size (width and height) of the LIDAR frame as shown in Figure 7. The LIDAR point cloud is converted into an elevation image in the same manner that is mentioned in Section 4.1. A histogram on the altitudinal differences between the pixels in the LIDAR and map elevation images is formed. The histogram is converted into a probability distribution vector $P_{obs}^t$ by normalizing the bars based on the Otsu threshold $l_t$ [24] and scaling up by the Sigmoid function as implied in Equation (13).

$$P_{obs}^t(\Delta z) = \left(1 + e^{-\alpha(l - l_t)}\right)^{-1}$$

(13)

![Figure 7. LSM based localization system in the Z plane. Elevation map image is cut off in the same size of LIDAR elevation frame (192 x 192) according to the obtained offsets in XY plane. Histogram between the elevation LIDAR and cut-map images is created and converted into a likelihood vector and then skinned by Sigmoid function to estimate z-offset based on Otsu’s threshold.](image)

A binary Bayes filter (BBF) is then incorporated to model the distribution variance with respect to the time [25] and calculate the altitudinal posterior probability as in Equation (10).

$$P_t(\Delta z) = \left(1 + e^{-L_t}\right)^{-1}, \quad L_t = \log\left(P_{obs}^t / 1 - P_{obs}^t\right) + \log\left(P_{obs}^t-1 / 1 - P_{obs}^t\right)$$

(14)
where $P_t$ is the posterior probability and $P_{t-1}$ is the prior probability. The position of the highest peak in the vector refers to the altitudinal offset that should be added to correct DR measurements and obtain the z vehicle position.

6. Experimental Results and Discussion

6.1. Setup Configuration

Figure 8 shows the experimental platform equipped with many sensors to enable autonomous driving. A Velodyne HDL-64E S2 with 64 laser beams is attached to the roof to scan the surrounding environments and generate 3D point clouds. Two GNSS antennas are also attached and connected to an Applanix POS-LV 220 (GIR) coupled GNSS/IMU system to record various measurements such as velocity and pose. These data are post-processed to produce a highly accurate vehicle trajectory in the ACS and the mapping module is then activated to generate elevation and intensity maps.

![Figure 8. Experimental platform with sensor configurations.](image)

During autonomous driving, the DR technique was employed to estimate the 3D vehicle position using the maps and the proposed localization modules in the XY and Z directions. In order to evaluate the system’s performance, the Applanix system was also activated to generate the ground truth. The central processing unit was deployed in the trunk and possesses Intel-Core™ i7-6700 CPU working at 3.40 GHz with 8 GB RAM. The operating system is Windows-7 64X and the localization and mapping systems were coded in VS-2010 C++ using OpenCV and OpenGL libraries.

The parameters of the LSM and localization systems are set as follows: $\lambda = 1024^2$ is the pixel number to create a node, the pixel resolution $\sigma$ is 12.5 cm in the real world and $\beta = 2$ m is the altitudinal threshold to divide ACS into layers. The size of a layer map sub-image is $S_w \times S_h = 512 \times 512$. In the localization module, the dimensions of the large map image is $2S_w \times 2S_h$ and the LIDAR image is in the size of $192 \times 192$. The maximum velocity during autonomous driving is set to 60 km/h.

6.2. Analysing the Mapping System in Ohashi Junction

As the main goal of this paper is to emphasize the robustness of the LSM, Ohashi Highway Junction in Tokyo was selected as shown in Figure 9. Ohashi Junction is one of...
the most complicated traffic nodes in Japan because it contains four stacked loops with a height of 71 m (36 m underground and 35 m of closed sky above ground). Each loop is 400 m lap in an upward/downward driving direction with two lanes of the width of 9.75 m. The maximum slope in the longitudinal direction is 7% whereas it is 10% in the lateral direction. The junction contains two entrances and two exits to drive through each loop in the up/down directions. On 9 October 2021, we drove from Yono Junction towards the Yamate Tunnel which ends with the underground entrance to Ohashi Junction. The upward two loops were then scanned and a U-turn was made shortly after the exit to scan the downward loops and collect the data of the GIR and LIDAR as illustrated in Figure 9.

**Figure 9.** Ohashi junction. The 35 m closed sky part aboveground with two driven upward and downward trajectories. The red dotted line demonstrates the vehicle trajectory in the forth loop.

Before using the LSM, the validation of the accumulation strategy had to be checked with respect to the vehicle trajectory. This was achieved by calculating the difference between the direct post-processed altitudinal vehicle position and the elevation value of the nearest pixel’s error on the map. Figure 10 shows the map-error profile of the trajectory between Yono Junction and the exit of Ohashi Junction. Along this trajectory, the road slope changes considerably at different segments, e.g., two bridges, the Yamate Tunnel and loops between 30 m underground and 35 m above ground. One can observe that the average error is less than 2 cm; this indicates the robustness of the proposed accumulation method as explained in Section 4. However, this reasonable error emerges because of the accuracy of the calibration parameters between the vehicle and LIDAR coordinate systems.

**Figure 11** demonstrates the continuous rapid change of the lateral and longitudinal slopes in the first and second upward loops. The images encode the top 200 cm altitude in the road and each bright-dark edge indicates an increment of the height by $\beta$. Saving these float changes in a dense pattern is vital to represent the road elevation accurately. Moreover, it is very important to consider the processing time for the map retrieval in such
a critical highway junction. These requirements are significantly achieved by the LSM by simply reading BMP images using the IDs \((z)\) to represent the height in the m-order and the pixel values to densely and continuously indicate the cm-order of the road surfaces.

Figure 10. The trajectory between Yono and Ohashi highway junctions with front camera images of ground, bridge, tunnel and entrance of Ohashi.

Figure 11 demonstrates the continuous rapid change of the lateral and longitudinal slopes in the first and second upward loops. The images encode the top 200 cm altitude in the road and each bright-dark edge indicates an increment of the height by \(\beta\). Saving these float changes in a dense pattern is vital to represent the road elevation accurately. Moreover, it is very important to consider the processing time for the map retrieval in such a critical highway junction. These requirements are significantly achieved by the LSM by simply reading BMP images using the IDs \((z)\) to represent the height in the m-order and the pixel values to densely and continuously indicate the cm-order of the road surfaces.

Figure 11. First and second loops in upward direction representing the top 200 cm of the road height. Bright-dark edges indicate increasing height by 2 m. Road slope is effectively and densely illustrated by grayness changes.

Figure 12a shows a sub-image of a common area of Ohashi Junction that illustrates the accumulation of the four loops with the exit. During autonomous driving, this inaccurate accumulated map image is compared with a LIDAR image of only a unique layer to localize the vehicle in the lateral and longitudinal directions. Therefore, achieving accurate localization is almost impossible and very risky because the map image contains different multiple matching patterns. In contrast, Figure 12b illustrates the efficiency of the LSM to automatically separate and encode the differences in the road texture and structure along the four loops using a set of layered images with the corresponding IDs in the XYZ plane.
Figure 12. Common area of four loops with the exit segment of Ohashi Junction. (a) Accumulative map image produced by the old mapping system [8]. (b) LSM with corresponding ID: elevation in 200 cm above Z-IDs (left) and intensity (right).
6.3. Localization in a Challenging Multilevel Environment Using LSM

In the proposed LSM-based localization framework, the accuracies in the XY and Z directions influence each other because $xy$ offsets are computed using a map image that is created based on the estimated altitudinal position $z$. The offset of $z$ is then calculated using the $xy$ estimated position. Figure 13 shows the error profiles of the elevation map and the estimated altitudinal position with respect to the ground truth in the upward and downward two loops, respectively. One can observe the high accuracy of the altitude estimation (around 2 cm). The position error is less than the map error that implies the robustness of the proposed method to relatively compensate the calibration errors between LIDAR and vehicle coordinate systems as well as to recover the slight changes in environments in the Z direction, such as grass on the roadsides. Accordingly, the map images in the four layers are perfectly retrieved and formed during autonomous driving as shown in Figure 13 for the common area in Figure 12b.

Figure 13. The elevation map error compared to GIR and the altitudinal accuracy in upward and downward trajectories with map images in the common areas (A–D) that were retrieved by the proposed localization system based on layered elevation-intensity images. The RMSE errors of the map and altitudinal errors in the upward and downward trajectories are 0.025 m, 0.019 m, 0.029 m and 0.011 m.

Figure 14 demonstrates the localization accuracy in the lateral and longitudinal directions of the estimated positions by DR and the proposed method in the upward and downward loops, respectively. The accuracy of the proposed method is less than 10 cm, whereas 30 m is the average drifting by DR. This indicates the robustness to perfectly
retrieve the road surface on each level and compensate the DR errors by a LSM. Accordingly, the scalability and applicability of the LSM and the relevant based localization system have been proved to precisely and safely enable autonomous driving in very critical multilevel environments.

Figure 14 demonstrates the localization accuracy in the lateral and longitudinal directions of the estimated positions by DR and the proposed method in the upward and downward loops, respectively. The accuracy of the proposed method is less than 10 cm, whereas 30 m is the average drifting by DR. This indicates the robustness to perfectly retrieve the road surface on each level and compensate the DR errors by a LSM. Accordingly, the scalability and applicability of the LSM and the relevant based localization system have been proved to precisely and safely enable autonomous driving in very critical multilevel environments.

Finally, one can claim that the GIR system may provide low-accuracy measurements in challenging environments because of the obstruction and deflection of the satellite signals, i.e., long tunnels and Ohashi Junction. This significantly influences the map creation and the layer separation. This problem has been robustly solved by proposing

![Graph showing localization accuracy](image-url)
a unique Graph-SLAM framework on the node level to generate precise nodes intensity and elevation images (2.5D maps) [26]. Consequently, the proposed LSM strategy in this paper was incorporated into the final stage of the Graph-SLAM framework to generate an accurate LSM.

7. Conclusions

We proposed a robust mapping framework to generate layered sub-image maps in multilevel environments using intensity and elevation LIDAR data. The unique strategy to use the 2D image domain to represent the elevation float values efficiently enabled map retrieval in real time by simply reading images during autonomous driving. This allowed us to encode a dense and continuous representation of the roads’ structures and slopes. Moreover, the identification tactic in the XYZ plane facilitated the update of the map on each layer and add new layers based on the vehicle trajectory. The reliability and efficiency of the proposed framework have been proved by generating the LSM in a very critical multilevel highway junction and accurately localizing the vehicle with average errors of 5 cm and 2 cm in the XY and Z planes, respectively.

Author Contributions: M.A.: Investigation, Methodology, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review and editing and Submission. N.S.: Supervision and Resourcing. R.Y.: Methodology, Investigation, data collection and resources. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the Japan Society for the Promotion of Science (KAKENHI Funds: No. 20K19893 and No. 22K17974).

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

References

15. Nam, T.H.; Shim, J.H.; Cho, Y.I. A 2.5D Map-Based mobile robot localization via cooperation of aerial and ground robots. Sensors 2017, 17, 2730. [CrossRef] [PubMed]
26. Aldibaja, M.; Suganuma, N. Graph SLAM-Based 2.5D LIDAR Mapping Module for Autonomous Vehicles. Remote Sens. 2021, 13, 5066. [CrossRef]