TRAINING DATA PRE-PROCESSING WITH MULTI-OTSU THRESHOLDING FOR IMPROVING ROAD MARKING EXTRACTION FROM SPARSE MOBILE LIDAR POINT CLOUD SCANNING-DERIVED IMAGES

Miguel Luis R. Lagahit\*1,2,3 and Masashi Matsuoka1,2

1 Department of Architecture and Building Engineering, Tokyo Institute of Technology, Japan

2 Tokyo Tech Academy for Super Smart Society, Tokyo Institute of Technology, Japan

3 Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Japan

Email: lagahit.m.aa@m.titech.ac.jp, matsuoka.m.ab@m.titech.ac.jp

**KEY WORDS:** Low-Cost Mobile Mapping, Road Marking Extraction, Convolutional Neural Networks

**ABSTRACT:** Many studies have successfully demonstrated the use of deep learning techniques to perform road marking extractions from dense point cloud-derived images obtained from survey-grade mobile mapping systems. Recently, in an effort to cut expenses, road marking extractions from sparse point cloud-derived images from low-cost LiDAR sensors, have also been explored. However, due to the resulting inadequate representation of features after scanning, road marking extractions using deep learning techniques posed to be challenging. To work around this issue, modifications of the various components of the deep learning framework have been investigated such as the loss function and the model structure itself. In this work, we adopt a different approach and consider including a pre-processing step when preparing the training dataset. Our proposed method considers two concepts introduced in previous research, Otsu thresholding and lidar value-derived image channels, by applying multi-Otsu thresholding to lidar point cloud-derived intensity images. In addition, we also examine the impacts of altering the number of channels that underwent thresholding and compare results with those of existing approaches. We utilized the Fast-SCNN model with Focal Combo Loss for the testing, which has already shown good performance in terms of both speed and accuracy in road marking extraction from sparse point cloud-derived images. Results reveal about 1.6% to 4.1% improvements in the resulting f1-score, depending on the number of channels thresholded, reaching around 71.7% to 74.1%. Considering that our proposed method deals with the training data, it could be easily introduced in existing and soon-to-be-developed deep-learning extraction methods.

# INTRODUCTION

## Background

A common procedure to automatically extract road marking features from rasterized dense point clouds is through the use of convolutional neural networks (CNN). Recently, to address the issue of HD map updating, the idea of utilizing low-cost sensors for practicality has been explored. This led to studies like that of (Lagahit and Matsuoka, 2023), that attempted to extract road marking features from sparse point clouds obtained during low-cost LiDAR scanning. However, this proved to be challenging since features on sparse point clouds are poorly represented and convolutional neural networks struggled in conditions with extreme class imbalance. (Lagahit and Matsuoka, 2023) proposed focal combo loss, to support the performance of CNNs in such cases of class imbalance and have proven to be successful even for lighter CNN models. Nonetheless, extractions still remain at a range of about 70% to 85% in terms of f1-scores, which signifies that there is still a huge room for improvement.

On the other hand, (Mattheuwsen and Vargauwen, 2020) and (Wu et al., 2020), have employed the use of variations in the representation of a rasterized point cloud to boost CNN segmentation performance. Such representation, has been coined the IHV representation, where I stands for intensity, H stands for height, and V stands for height variance. IHV, as the name suggests, makes use of different point cloud rasterization techniques for each of the image layers. Given their successful demonstration, that a change in rasterized point cloud representation contributes to better segmentations, it opens up the idea of exploring other methods of image layer representations.

In this paper, we follow through the ideas presented above and explore a new image presentation to improve road marking extraction on sparse point cloud-derived imagery. We propose the use of multi-Otsu thresholding, a variation of Otsu thresholding, that presents a method to initially segregate features into several classes, in preparing the training and testing datasets. The concept comes from the usage of Otsu thresholding in semi-automatic extraction workflows for HD map mapping (Chang et al., 2023). Intuitively, since it has demonstrated its usefulness in semi-automatic workflows it has a huge possibility to translate well for automatic workflows as well.

## Objective

This paper aims to explore the impacts of using multi-Otsu thresholding as a pre-processing step to enhance road marking extraction on sparse mobile LiDAR point cloud scanning-derived images. In doing so, we also provide (1) a comparison to the original dataset and to Otsu, (2) an analysis of the effects on various numbers of multi-otsu thresholded layers, and (3) a comparison with a previous work that presented a different representation of a rasterized point cloud.

# METHODOLOGY

## Dataset

The dataset used for this paper was requested from the works of (Lagahit and Matsuoka, 2023). It contains low-cost mobile LiDAR scanning point cloud-derived intensity images and their corresponding labels. Figure 2-1 shows a more detailed description of the dataset. The images on the dataset were obtained by projecting each individual LiDAR scan to a top-down 2D plane resulting in a size of 2048x512 pixels. As we can see from the Figure, the dataset poses two major challenges for road marking extraction: (1) the poor representation of the target feature on the sparse point cloud, and (3) the extreme class imbalance caused by the dominating number of black pixels with respect to the target pixels.

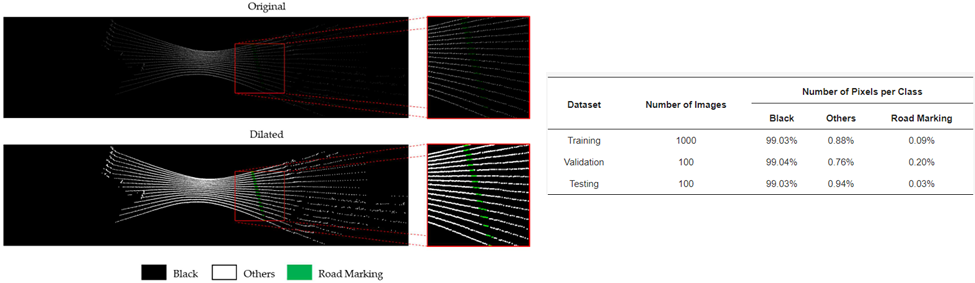


Figure 2‑1. (Left) Visualization of the images in the dataset, which is dilated for better visualization only, and (Right) the dataset statistics. Figures taken from (Lagahit and Matsuoka, 2023).

## Otsu and Multi-Otsu Thresholding

In this paper, we introduce multi-Otsu, a thresholding algorithm as a pre-processing step. Multi-Otsu is a variation of the Otsu algorithm that efficiently calculates several thresholds depending on the desired number of classes (Liao et al., 2001). In Figure 2-2, we can see the effects of the algorithms when applied to a grayscale image for generating a binary image. This will be useful, especially for isolating targets that have highly distinct pixel values that are small numbers as compared to the background, such as high-intensity road marking features on a rasterized point cloud, which will be shown in the following sections of this paper.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

Figure 2‑2. (Left) Histogram of the grayscale image and the thresholds, depicted by the red vertical line, calculated by (Center) Otsu and (Right) multi-Otsu. For the binary image depicted after multi-Otsu, we have set it as the furthermost section over all the others. The image containing Pikachu is taken from Pokemon.com and modified for example usage.

## Semantic Segmentation via CNN

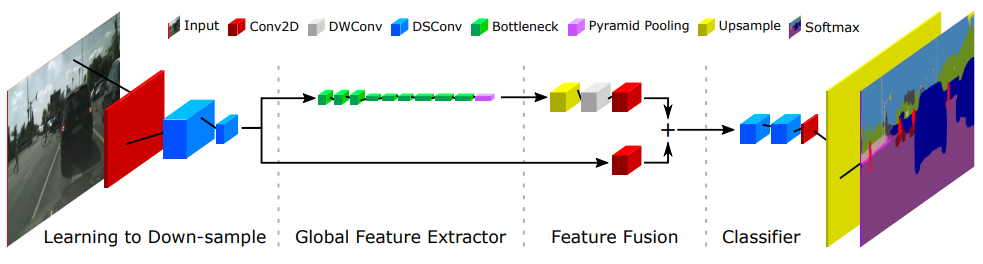


Figure 2‑3. The Fast-SCNN structure. Figure taken from (Poudel et al., 2019).

For the task of semantic segmentation, which categorizes every pixel in the image into a certain class, we employ the use of Convolutional Neural Networks (CNN). In this paper, we apply Fast-SCNN, a CNN model designed for real-time segmentation (Poudel et al., 2019). Not only is Fast-SCNN fast, but together with Focal Combo loss, it has proven to level with the results of U-Net, a popular CNN model used in various fields for semantic segmentation (Lagahit and Matsuoka, 2023). In addition, for replicability, the following set of parameters were used during training: a learning rate of 0.0001, a batch size of 16, and an ADAM optimizer. The images were also downscaled by a quarter to meet the processing capability of the computer.

## Assessment

To assess the impacts of adopting multi-Otsu thresholding as a pre-processing step, we make use of recall and precision. These are standard semantic segmentation metrics obtained from the confusion matrix. We also make use of the f1-score, which is the harmonic mean of precision and recall, as a final metric to provide a better evaluation of the performance in cases where the resulting precision and recall greatly differ. The Equations below denote the formulas to obtain the metrics from the confusion matrix.

|  |  |
| --- | --- |
| , | (1) |
| , | (2) |
|  | (3) |

In relation to the metrics given above, due to the properties of the dataset observed by (Lagahit and Matsuoka, 2013), we can remove misclassifications in the black pixel regions during computation. This is because pixels in those areas correspond to no point cloud value and thus contain no relevance to the segmentation task. This shift in perspective has been demonstrated to have a significant impact on how CNN’s performance on road marking extraction is evaluated. As can be seen in Figure 2-4, by employing this concept a significant increase in precision could be observed.

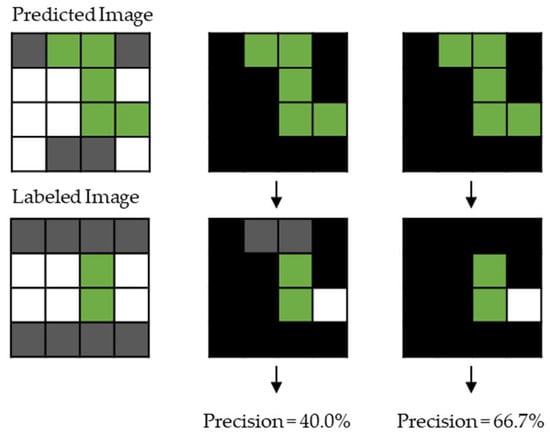


Figure 2‑4. A demonstration of the impacts on the precision value of omitting the misclassifications in the black region. Figure taken from (Lagahit and Matsuoka, 2023).

# Results and discussion

The segmented images presented from this point onward have been dilated by a 5x5 kernel to enhance visualization. Without dilation, it would be impossible to observe the changes. Figure 3-1 depicts sample corresponding image pairs of an intensity image and its label, which can be used as the ground truth and reference for the succeeding results.

|  |  |
| --- | --- |
|  |  |

Figure 3‑1. Sample (Left) intensity image and (Right) its label from the testing dataset.

## Training Dataset Pre-Processing with Multi-Otsu Thresholding

When compared with Otsu, we can better segregate pixels that have high-intensity values by using multi-Otsu, where the number of desired classes is set to four. As an example, we can observe the histogram of a sample image in the testing data as illustrated in Figure 3-2. We can see that multi-Otsu successfully differentiates the small number of pixels with high intensity from the vast majority of pixels with values zero and close to zero. As such, for this paper, we select the fourth class over the others in producing the binary image.

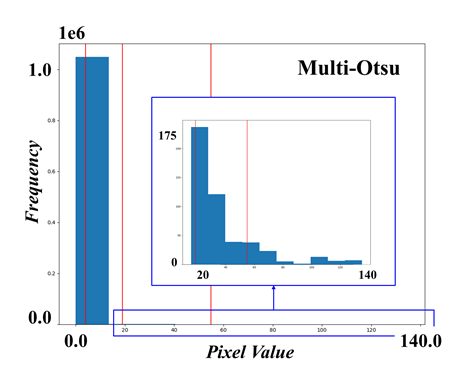
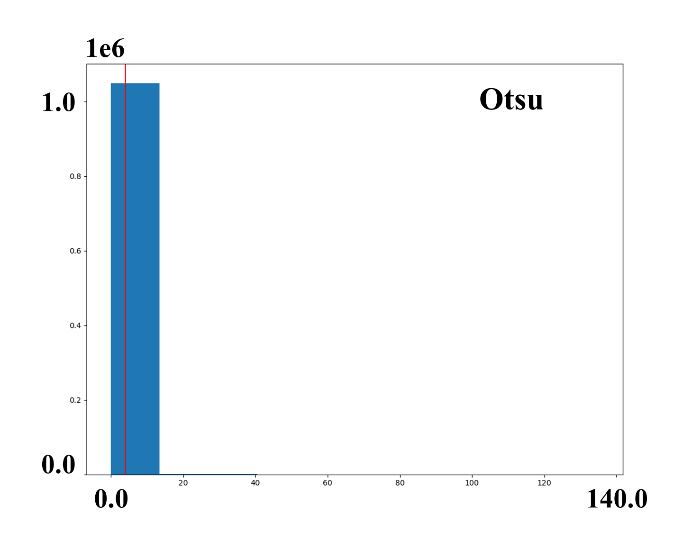


Figure 3‑2. Sample histogram of an intensity image from the testing dataset. The red vertical lines indicate the thresholding points for the image.

Figure 3-3 and Table 3-1 compare the road marking extraction results of Fast-SCNN with focal combo loss on datasets pre-processed with Otsu and multi-Otsu thresholding to the original datasets. Looking at the thresholded intensity images, it is clear that it largely helps in isolating our target road marking features. However, it still proves insufficient because other high-intensity features (e.g. vegetation) in or near the roadway are still present. Focusing on the resulting segmentations of our CNN, we can observe that by using the multi-Otsu pre-processed dataset, it was able to detect road marking features better than the original. Moreover, since Otsu failed to serve as an initial segregation procedure, the segmentation done by CNN also went poorly. This is reflected in our numerical assessment where multi-Otsu caused an increase of roughly 4% and Otsu caused deterioration of roughly 57% in their f1-scores. The increase was largely due to the significant increase in recall, which is a positive for mapping because it reduces the frequency with which it fails to detect road markings. However, we could see a drop in precision as a result of its poorer representation of the target geometry, meaning that surrounding pixels are also misclassified as the target.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Figure 3‑3. Sample segmentation results. From top to bottom: original, Otsu thresholded, and multi-Otsu thresholded. From Left to Right, the input intensity image, segmentation results, and the segmented road marking class that have been projected onto pixels with point cloud value (This orientation will be the same for the succeeding figures depicting segmentation results as well, so to remove redundancy it will be removed in the following captions.).

Table 3‑1. The assessment results of using multi-Otsu thresholding.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Recall** | **Precision** | **F1-Score** |
| Original | 64.35 | 76.92 | 70.08 |
| Otsu Threshold | 15.87 | 10.58 | 12.70 |
| Multi-Otsu Threshold | 80.46 | 68.72 | 74.13 |

## Different Levels of Multi-Otsu Threshold Image Layers

Since adapting multi-Otsu thresholding to the dataset has proven to be beneficial, we further investigate its performance when it is not fully applied to every layer. The results of multi-Otsu thresholding adapted to a single or two layers of an image are shown in Figure 2-6 and Table 3-3. We can observe that as we decrease the layers that undergo multi-Otsu thresholding so does its resulting f1-score. However, we can see that resulting f1-scores from thresholding two layers and all layers are quite near each other, with the prior having better recall and the latter having better precision. Most importantly, all of them outperformed the resulting f1-score of the CNN trained on the original dataset.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Figure 3‑4. Sample segmentation results. From top to bottom: multi-Otsu applied to all layers, two layers, and one layer.

Table 3‑2. The assessment results of using multi-Otsu thresholding at varying number of image layers.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Recall** | **Precision** | **F1-Score** |
| Multi-Otsu Threshold | 80.46 | 68.72 | 74.13 |
| IOO Representation | 83.36 | 65.55 | 73.39 |
| IIO Representation | 73.24 | 70.14 | 71.66 |

## Comparing with the IHV Representation

Finally, we compare our results to a previous work that applied different point-cloud-to-image approaches to each individual image layer, namely the IHV representation. IHV was introduced by (Mattheuwsen and Vergauwen, 2020) to introduce a new representation of a rasterized point cloud, wherein I is for intensity, H is for (minimum) height, and V is for (height) variance. They have shown that using this representation produced good results in terms of manhole extraction using a CNN. Unfortunately, as seen in Figure 3-7 and Table 3-4, this method fails for our purpose of road marking extraction. Both recall and precision went down resulting in a 74% decrease in the f1-score.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

Figure 3‑5. Sample segmentation results. (Top) multi-Otsu thresholded intensity image and (Bottom) IHV representation.

Table 3‑3. The assessment results of the proposed multi-Otsu thresholding in comparison to the IHV representation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Recall** | **Precision** | **F1-Score** |
| Multi-Otsu Threshold | 80.46 | 68.72 | 74.13 |
| IHV Representation | 23.10 | 2.27 | 4.14 |

To further analyze the failure of the IHV representation, we examine the results when H and V are used to replace a single layer in an intensity image. From Figure 3-8 and Table 3-5, we can see that both of them failed, resulting in large misclassified regions and none of the target features extracted. So, it is not surprising that the IHV representation fails because its components drastically degrade the segmentation performance of the CNN. This can be related to our target road markings, which, unlike manhole covers, lie flat with the ground surface, so point-cloud properties with regard to height may not be as relevant.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

Figure 3‑6. Sample segmentation results. Using intensity image, substituting one layer with (Top) minimum height image and (Bottom) height variance.

Table 3‑4. The assessment results of substituting one layer of the intensity image with H and V.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Recall** | **Precision** | **F1-Score** |
| HII Representation | 6.49 | 9.20 | 7.61 |
| VII Representation | 34.33 | 2.80 | 5.18 |

# CONCLUSION

This paper attempted to explore the effects of utilizing multi-Otsu thresholding as a pre-processing step to improve road marking extraction on sparse mobile LiDAR point cloud scanning-derived images. We were able to demonstrate a maximum of around 4% increase in f1-score as compared to the model trained using the original dataset and an increase of 70% as compared to the previous work (IHV). Further ablation studies on a larger dataset shall be done in the future to confirm the validity of the proposed multi-Otsu pre-processing.

# References

Chang, Y.F., Chiang, K.W., Tsai, M.L., Lee, P.L., Zeng, J.C., El-Sheimy, N., Darweesh, H., 2023. The Implementation of Semi-Automated Road Surface Markings Extraction Schemes Utilizing Mobile Laser Scanned Point Clouds for HD Maps Production*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 93-100.

Lagahit, M.L.R., Matsuoka, M. 2023. Focal Combo Loss for Improved Road Marking Extraction of Sparse Mobile LiDAR Scanning Point Cloud Derived Images using Convolutional Neural Networks. *Remote Sensing*, 15, 597.

Liao, P.S., Chen, T.S., Chung, P.C., 2001. A Fast Algorithm for Multilevel Thresholding. *Journal of Information Science and Engineering*, pp. 713-727.

Mattheuwsen, L., Vergauwen, M. 2020. Manhole Cover Detection on Rasterized Mobile Mapping Point Cloud Data using Transfer Learned Fully Convolutional Neural Networks. *Remote Sensing*, 12, 3820.

Poudel, R., Liwicki, S., Cipolla, R. 2019. Fast-SCNN: Fast Semantic Segmentation Network. *arXiv.*

Wu, H., Xie, Z., Wen, C., Wang, C., Li, J., 2020. On-Road Information Extraction from LIDAR Data via Multiple Feature Maps*. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 207-213.