Road Detection Method based on Online Learning

Wenbo Wang[1]

Nanjing Audit University

NanJing, China 211899

Yong Ma[2]

Nanjing University of Science and Technology

NanJing, China 210014

Abstract—**Road detection is always the key problem of researches on areas of unmanned ground vehicle and computer vision. A road detection method is proposed based on online learning and multi-sensor fusion. First of all, the Lidar point clouds are projected onto the images via the joint calibration of these two kinds of sensors. Then Simple Linear Iterative Clustering is used to segment images into many superpixles. Based on that, a multilayer online learning method is proposed, in which 2 Support Vector Machines are trained to detect the road. To be specific, the superpixel layer Support Vector Machine is used to detect road roughly, and the pixel layer Support Vector Machine is then trained to classify the edge pixels of the road areas, which is classified by the upper-layer Support Vector Machine. These 2 Support Vector Machines are updated online at each frame to be adapted to the changing environment. At last, some experiments are carried out on KITTI RAW dataset and an autonomous land vehicle, and the results show the effectiveness of proposed method. The main contributions of this work lie on as follows: 1) a multilayer learning model is proposed to detect road more robustly and accurately; 2) an online learning method is proposed which can be adapted to the changing environment.**

Keywords—*Road detection; data fusion; unmanned ground vehicle; online learning; image segmentation*

I. Introduction

Road detection [1], [2], [3] is one of the key technologies in multiple research areas such as unmanned ground vehicle development and machine vision. Traditional road detection methods are based on image data captured by RGB cameras mostly [4]. In the past decades, researchers have provided bunches of image processing algorithms [5], [6], [7], such as vanishing point localization, natural road boundaries detection or CRF [8] (conditional random fields) optimization [9] after segmentation of pixels based on prior information. However, different illumination conditions, shadow and complicated texture background significantly impact on image quality, which leads to rapid decline of algorithm performance. Recently, with the development of 3 dimensional sensor, researchers have put forward various road detection methods based on range data. For example, Sunando Sengupta and Paul Sturgess constructed an octree model to describe the detected environment and attached an advanced CRF model for semantic segmentation [10]. Benjamin Suger published a semi-supervised machine learning method, using Lidar to construct an accessible probability map, for outdoor navigation of robots [11].

Compared to RGB cameras, three dimensional sensors outperforms in many areas. Range data of surrounding is detected from all directions based on 3D sensors [12], [13], providing adequate information of target structure without interference from illumination conditions, severe shadow, complicated texture background and etc. However, 3D sensors have

their own weaknesses. First, taking stereoscopic vision into account such as Kinect, they are influenced easily by moving targets, leading to large amount of noise in detected range image [14], [15]. Furthermore, their observation range is more than 20 meters usually, which could not fulfil the requirements of unmanned ground vehicle environment perception. Second, although the observation range of Lidar is 120 meter in maximum, the detecting data is increasingly sparse with the increasing distance, due to its fixed angular resolution on measuring targets. So, Lidar senor could not describe complete terrain or target detail appearance in distance [16], [17].

In order to combine the advantages of these two kinds of sensors, increasing number of researchers fuse the outcomes of these two to acquire better environment model, especially on road detection [18], [19]. In addition, due to the continuous surrounding changes of unmanned ground vehicle, it is difficult for the pre-trained classifier to perform well on road detection and classification. In order to deal with this problem, unmanned ground vehicles are required to do online learning [20], [21] based on real-time surround environment. In this paper, a hierarchical online learning method is put forward and verified its efficiency on KITTI raw dataset and our own unmanned ground vehicle. The main innovation of this paper is as follows. (1) A hierarchical model is introduced, which could realize robust road detection more accurate. (2) An online learning method is put forward, which is adaptive to continuous environment changes.

The main content of this paper is organized as follows: In Section II, we briefly review the fusion of image and lidar data. In Section III, we will propose our model in detail. In Section IV, several experiments are designed to verify the effectiveness of our model. In Section V, some conclusions are given to finish this paper.

II. Fusion of Image and Lidar Data

To combine image and range data, joint calibration is needed for data alignment of camera and Lidar. Lidar point cloud is expressed as $\mathbf{P}_{lidar} = \{\mathbf{X,Y,Z}\}$ and its corresponding projection outcome is recorded as $\boldsymbol{P}_{image} = \{\boldsymbol{U,V}\}$

$$\boldsymbol{P}_{\text{image}} = \boldsymbol{R}_{\text{rect}}^0 \boldsymbol{T}_{\text{lidar}}^{\text{image}} \boldsymbol{P}_{\text{lidar}}$$
$$\boldsymbol{T}_{\text{lidar}}^{\text{image}} = \begin{pmatrix} \boldsymbol{R}_{\text{lidar}}^{\text{image}} & \boldsymbol{t}_{\text{lidar}}^{\text{image}} \\ 0 & 1 \end{pmatrix} \tag{1}$$

In which, $\boldsymbol{R}_{rect}^0$ R is the matrix to transform original image visual angle to front view, while image lidar $\boldsymbol{T}_{lidar}^{image}$ is the matrix to project Lidar point to image view. $\boldsymbol{R}_{lidar}^{image}$ and $\boldsymbol{t}_{lidar}^{image}$ are rotation matrix and translation vector to project Lidar point
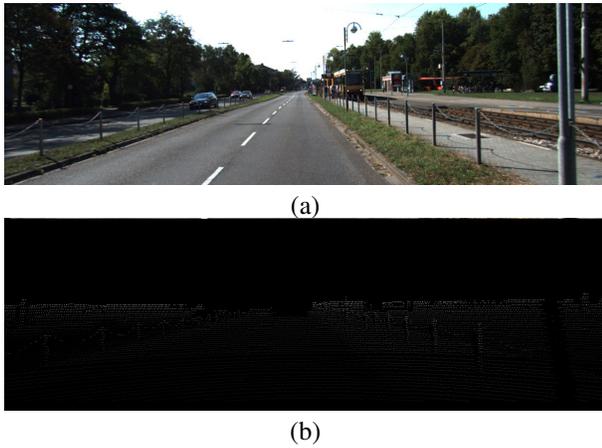
(a)

(b)

Fig. 1. Result of Lidar Point Clouds Projection. Subfigure (a) is the Original Image Captured by RGB Camera, while Subfigure (b) is the Lidar Projection Outcome.

(a)

(b)

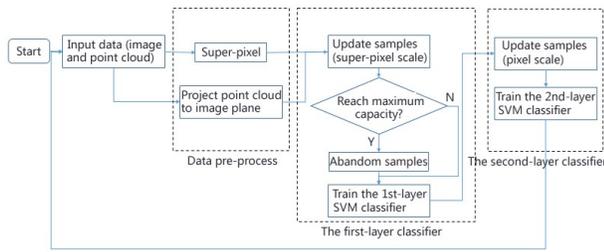Fig. 3. Result of Super-pixel Segmentation based on SLIC.

Fig. 2. Multiple Layers Online Learning Model.

cloud to image layer respectively. In many accessible datasets, such as KITTI dataset, calibration parameters are in open access, so we skip the detail computation procedure of these parameters. Details of joint calibration computation could be found in reference [22].

After joint calibration mentioned above, Lidar point cloud is projected to image plane, as Fig. 1 shown. Subfigure (a) is the original image captured by RGB camera, while subfigure (b) is the Lidar projection outcome, in which height feature is indicated by brightness. With joint calibration and projection procedures, some pixels on image is corresponding to Lidar point, which means their range and height is available.

III. MULTIPLE LAYERS ONLINE LEARNING MODEL

A. Hierarchical Online Learning Definition

In order to adapt to continuous changing background environment in road detection, we put forward a multiple layers online learning model as Fig. 2 shown.

In the rest of this paper, we will introduce this model in details.

B. Super-pixel and Road Detection Definition

Super-pixel is a series of adjacent pixels composed of small areas with similar color, brightness and texture characteristics. Most of these small areas retain the effective information for further image segmentation, and generally do not destroy the
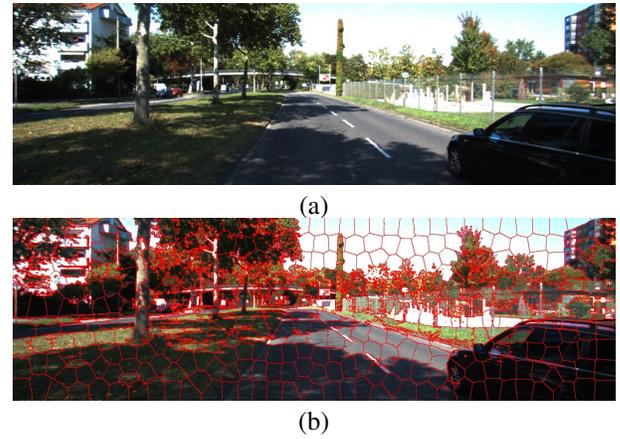
boundary information of objects in the image. Therefore, more and more image segmentation algorithms adopt super-pixel as the basic segmentation unit [21], [23]. As Fig. 3 shown, SLIC super-pixel method is utilized to several super-pixel units in this paper. In super-pixel procedure, segmentation is expressed as $R = \{r_i, i=1..M\}$, in which M is the number of super-pixels. Then, road detection can be considered as a bi-classification problem:

$$
\begin{aligned}
r_i &= (F_i, L_i) \\
F_i &= \{f_{i1}, f_{i2}, \ldots, f_{id}\} \\
L_i &= \{+1, -1\}
\end{aligned}
\tag{2}
$$

Here, F_i is the corresponding characteristic of r_i, d is the feature dimension and L_i represents the final classification outcome, in which L_i=+1 indicates that rF_i belongs to road region, and L_i=-1 means that r_i is an off-road region.

C. Choice of Classifier

In recent years, deep learning and other new machine learning methods have been extensively studied, and have achieved remarkable results on many datasets. However, deep learning requires a large number of training samples to learn network parameters, and requires very high computational resources. Even with the use of transfer learning and other learning technologies, the deep learning method is difficult to apply in the specific task scenario of online road detection. Compared with deep learning method, the final decision function of Support Vector Machine (SVM)[24] classifier is determined only by a few support vectors rather than all training samples. Therefore, the computational complexity is low and key samples can be calculated automatically to realize efficient machine learning. In the real-life scene using online road learning, the sample size is often small, and the distribution of positive and negative samples is very uneven in the initial stage of learning. Therefore, compared with other classifiers, SVM has excellent performance in online learning system. In addition, in view of the problem that the computational complexity of SVM classifier will raise with the increase of training samples, the online learning model proposed in this paper sets a maximum sample size and an update strategy
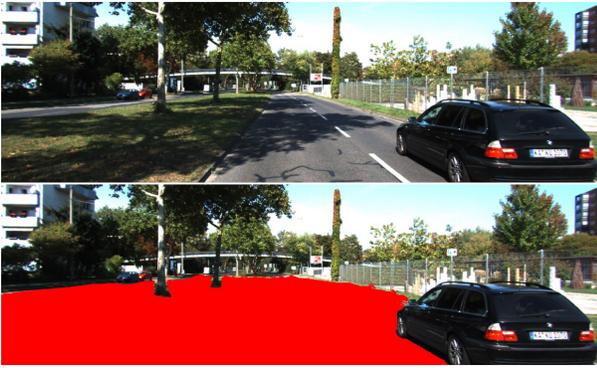
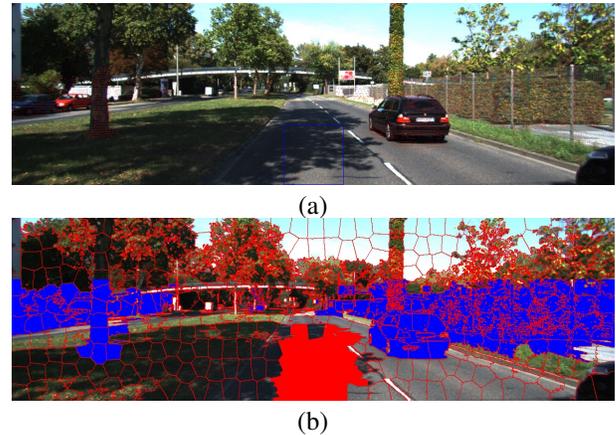Fig. 4. Road Detection Result Only via Lidar Data.

(a)

(b)

Fig. 5. Selection of Training Samples of Classifier in the First Layer.

of positive and negative samples to ensure that the proposed method can meet the real-time requirements of road detection.

D. Road Detection in Super-pixel Level

In the multi-level learning model proposed in this paper, the first level classifier is at super-pixel scale. Generally speaking, one of the important factors affecting the performance of online road detection methods is the selection strategy of positive and negative samples. Many image-based online road detection algorithms assume a small area at the middle-bottom of the image as an initial road area and the image edge belonging to a non-road area [9], [21] when selecting classifier samples. However, in the actual scene, the edge of the image is mostly sky or building, which can not represent all kinds of negative samples near road. On the contrary, in many road detection methods based on Lidar, there are some problems in extracting positive samples. A typical algorithm for road detection based on Lidar data is to project the point cloud data of Lidar into two-dimensional grid map, then calculate the height difference of each grid. By setting a threshold artificially, the grid whose height difference is lower than the threshold is taken as a positive sample and the other grid as a negative sample [25]. As shown in Fig. 4, the first picture is the original image captured by RGB camera, and the area labelled in red in the second image indicates the positive sample area. In such methods, off-road regions with small height difference (such as lawn and road) can not be effectively distinguished.

In conclusion, image data is more suitable for extracting positive samples, while Lidar data is more suitable for extracting negative ones. Therefore, in this paper, we use a fusion method in this paper to combine Lidar and image data. Specifically, we assume that the super-pixels in the rectangular frame (as shown in Fig. 5(a)) at the image bottom are positive samples. Then use the Lidar data to separate obstacles from non-obstacles, and assume that the super-pixels belonging to obstacles are negative samples.

In order to figure out the super-pixels which belong to obstacles, Lidar point cloud data is projected to a plane constructed by X and Y coordinate axis. The plane is described in grid map form, then the maximum height difference of each grid is calculated and an artificial threshold is set to separate road girds and obstacle grids. Afterwards, project all obstacle grids to image plane as shown in red pixels in Fig.

5(a). Finally, those super-pixels with over a threshold of Lidar obstacle pixels are labelled as negative samples. As can be seen in Fig. 5(b) positive and negative samples are colored in red and blue respectively.

For each super-pixel unit, the color histogram and LBP texture features in HSI space are extracted from the color image, and the average height feature and height variance feature are extracted from the projected Lidar point image. After that, these samples are added to the sample library, and the first-level SVM classifier is updated and trained online. Then, all the super-pixel units in the whole image are classified by this classifier, and the first-level road detection results are obtained.

In addition, the online learning model sets the total capacity of the training sample library when updating the first level classifier online for each frame. If the current training sample library is not full, the positive and negative samples of this frame are directly added to the training sample library. If the current training sample library capacity has reached the maximum, then according to the proportion of positive and negative samples in the current sample library and sample collection time, part of the samples are deleted to make room for samples selected from current frame. Specifically, if there are more positive samples than negative samples in the current sample bank, the positive samples will be deleted, and on the other hand, the negative samples will be deleted. Secondly, the oldest samples will be deleted first to ensure that the trained classifier can continuously adapt to the latest environment.

E. Road Boundary Classification in Pixel Scale

In the super-pixel-based road detection algorithm, different scale setting of super-pixels has a significant impact on the accuracy of road detection. Many researchers are trying to solve this problem by using the idea of stratification. For example, document [21] proposes a multi-scale learning framework. They set up θ(odd) super-pixel segmentation layers of different scales from small to large, then trained a SVM classifier at each level, and used a voting method to get the final classification results next. However, in our research procedure, scale of super-pixels only influence road boundary pixels, while inner road regions are not sensitive to super-pixel scale as can be

seen in Fig. 6. In each image, super-pixel segmentation is shown in upper half, while road detection outcome in the below half. In that, this kind of methods witness a narrow promotion after consuming large amount of computation resources. To reduce the complexity of algorithm as well as ensure the accuracy of boundary localization, we train a second layer of SVM classifier in pixel scale based on road boundary super-pixels to polish up road detection result, after acquiring the first layer segmentation procedure. To be specific, we suppose road super-pixel set as R_{road}, which is segmented by the first layer classifier, while E_i indicates whether ith super-pixel belongs to road boundary, and its neighbourhood is expressed as η_i, Then E_i is calculated as follow.

$$E_i = \begin{cases} 1, & \text{if } \exists r_j \notin R_{road} \& r_j \in \eta_i \\ 0, & \text{else} \end{cases} \tag{3}$$

In the second SVM classifier layer, all pixels in road boundary super-pixels ($E_i = 1$) are become samples to be classified. Here, all pixels on road ($E_i = 0$) are used as positive samples, while pixels in obstacle super-pixels in the first layer are used as negative samples.

Unlike the updating method of the first-layer classifier, the second-layer classifier does not retain the samples of historical frames, but extracts the RGB values of all positive and negative samples of the current frame, as well as the average height and height variance of the super-pixels as features from each frame. We retrain the classifier in this layer, and segments all samples to be classified. Fig. 7 shows an example of sample choosing for the second-layer classifier. In Fig. 7, red and blue region indicates all the positive and negative sample pixels, respectively to train the second-layer SVM classifier, while the green region indicates all the pixels to be classified (road boundary region).

Finally, after the first-layer super-pixel classifying, the second-layer of our method separate road and off-road pixels in road boundary regions (green areas in Fig. 7). Combining output of the first-layer, the final road detection result is observed. The efficiency of the edge polishing up procedure is proved by experiments afterwards.

IV. EXPERIMENTS

To verify the efficiency of proposed algorithm, we choose a dataset (2011_09_26_drive_0013) randomly from KITTI RAW DATA. This dataset contains 143 continuous frames with a resolution of 1242×375 pixels, as well as corresponding Lidar frame with over 100 thousand points and the joint calibration parameters. The KITTI RAW DATA does not provide ground truth, so we label road regions on each frame artificially for algorithm verification.

Fig. 8 shows the number of training samples of the first-layer classifier with time accumulating. As can be seen in Fig. 8, positive samples are less than negative ones at the very beginning in the super-pixel classifier layer with a training bank containing 5000 training samples in total. However, with time going, the difference between positive and negative sample number starts to decrease after the 47th frame. The oldest sample keeps to be replaced continuously, and reaches
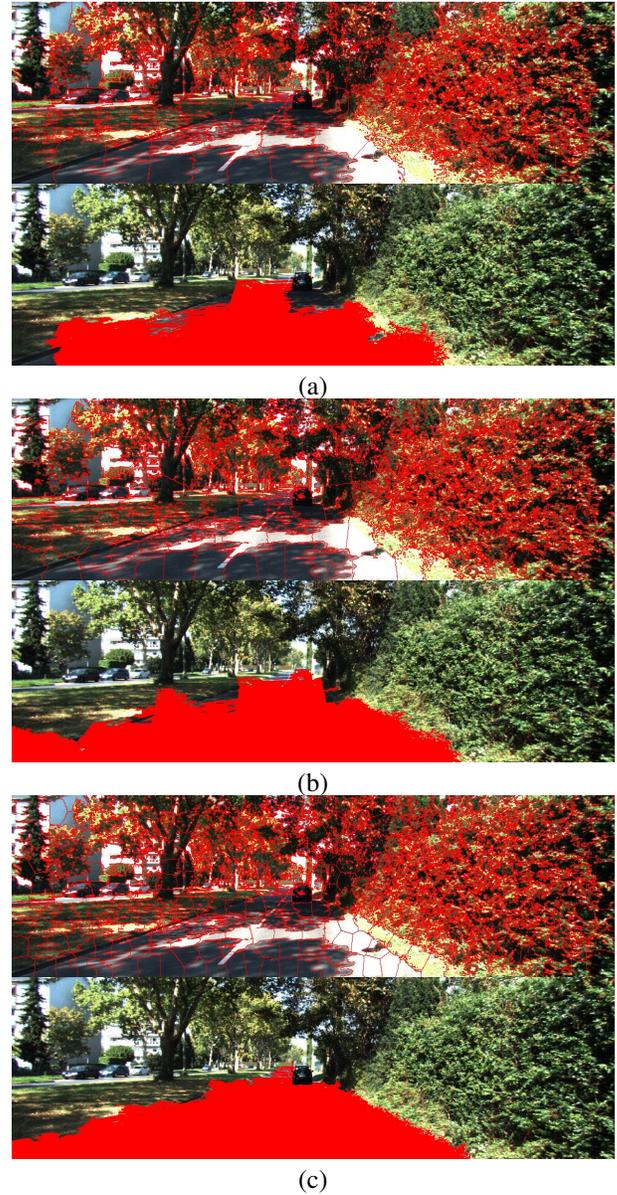

(a)

(b)

(c)

Fig. 6. Road Detection Results in Super-pixel Layer of Different Scales. Super-pixel Number is Set at 100 in (a), 200 in (b) and 500 in (c).
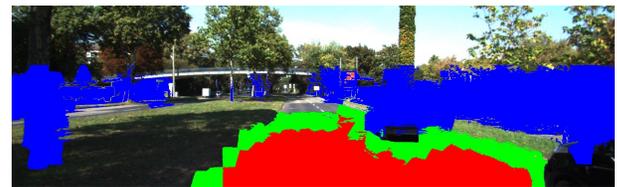

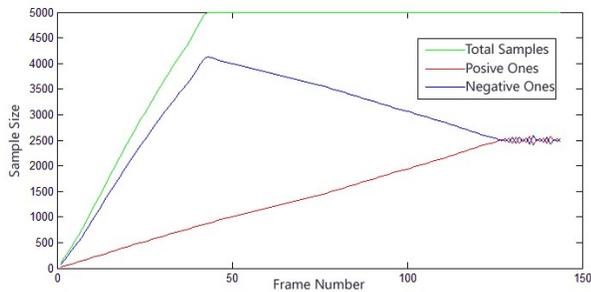Fig. 7. Selection of Samples of Classifier in Pixel Layer.

Fig. 8. Number of Training Samples of the SVM Classifier in the First Layer.

a balance with almost the same percentage. An uneven classi-fying problem becomes an even one at this point.

Fig. 9 illustrates road detection results of single layer (first layer) and multiple layers (total structure) learning models. As can be seen in Fig. 9, multiple layers model outperforms the single layer model. Due to an independent classification procedure of road boundary pixels, it can make full use of the road boundary information and detect road border well, which lead to higher road detection accuracy.

In order to further verify the efficiency of multi-sensors fusion and hierarchical online learning model, we realize four kinds of road detection methods on KITTI RAW DATA:

(1) Compute height difference in each super-pixel using Lidar data [25], then set a threshold (25 cm) artificially. Label super-pixels with height difference lower than this threshold as road regions, while the rest as non-road regions.

(2) Attach the method proposed in document [21] and use multi-scale super-pixel voting model for road detection.

(3) Use the fusion method proposed in this paper, by extracting positive and negative sample. Then detect road combining the method proposed in document [21], which uses multi-scale super-pixel voting model to detect road.

(4) Use the proposed multi-sensors fusion method in this paper to extract positive and negative samples, then detect road by multi-layers online learning model.

In this paper, we verify these four kinds of online road detection models' efficiency by six different parameters: FPR, TPR, Precision, Recall, Accuracy and F-measure. Table I put forward all kinds of parameters of these four learning models.

As can be seen in Table I, model (3) witness a significant promotion on road detection efficiency compared to model (1) and (2), so that the multi-sensors fusion efficiency can be proved compared to the single sensor road detection. In addition, Model (4) achieves the best performance, which further proves that multi-layer online learning model maks efforts to better road detection.

Next, we test the proposed method on our own unmanned ground vehicle to verify the efficiency in actual driving sce-nario. The unmanned ground vehicle is shown in Fig. 10. This vehicle contains a pre-calibrated RGB camera, Velodyne HDL64 Lidar and other kinds of sensors.

In our experiments, 3000 frames are collected in total as a dataset. Each frame contains a RGB image and corresponding
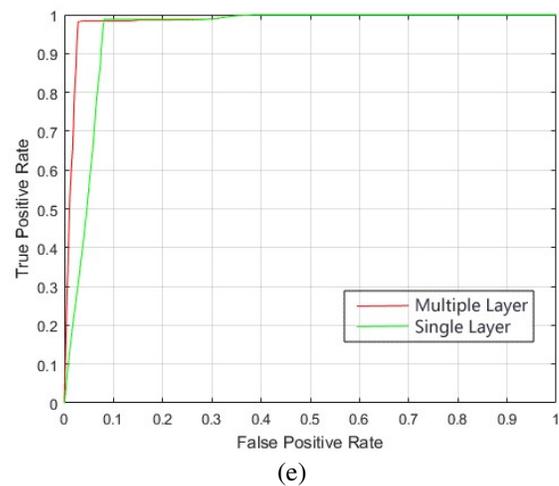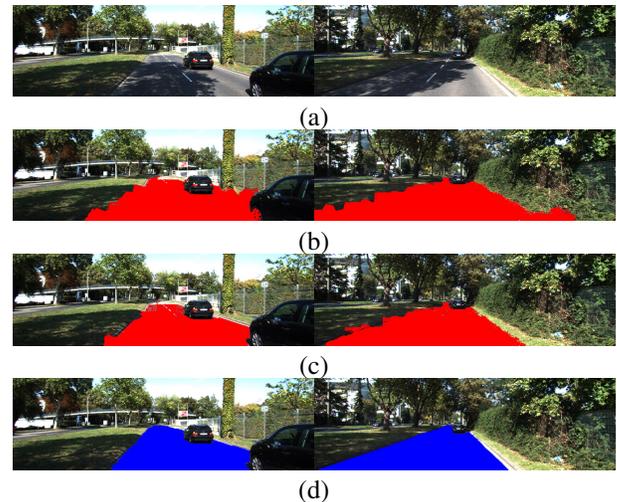

(a)

(b)

(c)

(d)

(e)

Fig. 9. Road Detection Results in Super-pixel Layer of Different Scales. (a) illustrates the Original Image of Two Input Scenario, (b) Shows the Outcome of the First-layer, (c) Indicates the Outcome of the Second-layer, and (d) is the Ground Truth, while (e) is the ROC Curve of Single Layer and Multiple Layers Models.

TABLE I. PERFORMANCES OF DIFFERENT ROAD DETECTION METHODS

Model	FPR	TPR	Precision	Recall	Accuracy	F-measure
Model (1)	0.0846	0.8043	0.6517	0.8043	0.8161	0.7593
Model (2)	0.0740	0.9301	0.7033	0.9301	0.9241	0.7941
Model (3)	0.0734	0.9313	0.7126	0.9313	0.9276	0.8033
Model (4)	**0.0680**	**0.9466**	**0.7178**	**0.9466**	**0.9343**	**0.8165**

Lidar point cloud data, as well as ground truth with artificial labelled road region. Finally, the proposed method achieves 91.07% precision on this dataset with an average running time at 87.35ms per frame, which meets real-time requirements. Fig. 11 shows part of our road detection results.

V. CONCLUSION

This paper proposes a road detection model used on unmanned ground vehicle based on online learning and multi-sensors fusion. According to our model, SLIC method is first utilized to separate image data to several super-pixels. Then Lidar data which belongs to obstacles is projected to image

Fig. 10. An Autonomous Land Vehicle.
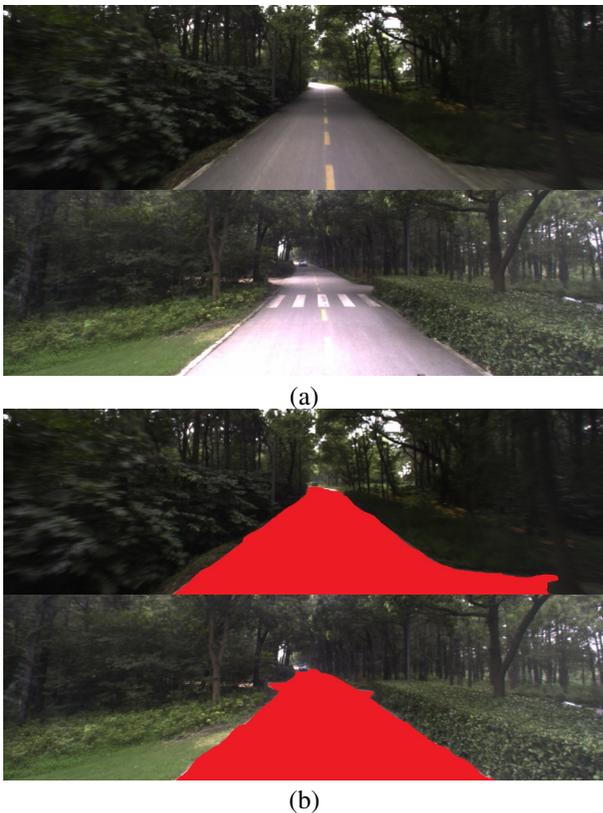

(a)

(b)

Fig. 11. Some Experimental Results. (a) Proposes Two Input Images, while (b) Indicates the Corresponding Road Detection Outcomes.

plane to separate image super-pixels into two kinds: obstacles and non-obstacles. Here in our method, we assume that super-pixels at mid-bottom of the image belongs to road region.

Afterwards, we put forward a multi-layer online learning model. In the first layer, large scale road detection is fulfilled by a SVM classifier, which trained by road and obstacle super-pixels. Next, another SVM classifier is developed for meticulous road detection in boundary regions. We utilize a new strategy to update the training sample bank, which could balance the percentage of positive and negative samples auto-matically. Maximum sample amount is limited to deal with the distribution problem of training data. Real-time requirements are met, while hierarchical classifier online learning is also accomplished to adapt to the environment changes.

The experiments performed on KITTI RAW DATA and our unmanned ground vehicle confirm that the proposed method meets real-time requirements in online learning road detection.

REFERENCES

[1] B. S. Shin, Z. Xu, and R. Klette, "Visual lane analysis and higher-order tasks: a concise review," *Machine Vision & Applications*, vol. 25, no. 6, pp. 1519–1547, 2014.

[2] V. Castillo, A. Díaz-Lvarez, M. Callejo, and F. Serradilla, "Grammar guided genetic programming for network architecture search and road detection on aerial orthophotography," *Applied Sciences*, vol. 2020, no. 10, 2020.

[3] T. Almeida, B. Loureno, and V. Santos, "Road detection based on simultaneous deep learning approaches," *Robotics and Autonomous Systems*, vol. 133, p. 103605, 2020.

[4] W. U. Hua-Yue and L. R. Duan, "Unstructured road detection method based on rgb entropy and improved region growing," *Journal of Jilin University(Engineering and Technology Edition)*, 2019.

[5] H. Kong, J. Y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 19, no. 8, p. 2211, 2010.

[6] J. M. Alvarez, A. M. Lopez, T. Gevers, and F. Lumbreras, "Combining priors, appearance, and context for road detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1168–1178, 2014.

[7] C. F. Ndez, R. Izquierdo, D. F. Llorca, and M. A. Sotelo, "Road curb and lanes detection for autonomous driving on urban scenarios," in *IEEE International Conference on Intelligent Transportation Systems*, 2014.

[8] J. Lafferty, A. Mccallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th International Conf. on Machine Learning*, 2001.

[9] K. Lu, L. Jian, X. An, and H. He, "A hierarchical approach for road detection," in *IEEE International Conference on Robotics & Automation*, 2014.

[10] S. Sengupta and P. Sturgess, "Semantic octree: Unifying recognition, reconstruction and representation via an octree constrained higher order mrf," in *IEEE ICRA*, 2015.

[11] B. Suger, B. Steder, and W. Burgard, "Traversability analysis for mobile robots in outdoor environments: A semi-supervised learning approach based on 3d-lidar data," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015, pp. 3941–3946, 2015.

[12] K. Rao and J. He, "Analysis on defect characteristics of urban road detection based on 3d ground penetrating radar," *Urbanism and Architecture*, 2019.

[13] S. Gu, T. Lu, Y. Zhang, A. Jose, J. Yang, and H. Kong, "3d lidar + monocular camera: an inverse-depth induced fusion framework for urban road detection," *IEEE Transactions on Intelligent Vehicles*, vol. PP, no. 3, pp. 1–1, 2018.

[14] J. Smisek, M. Jancosek, and T. Pajdla, "3d with kinect," *Advances in Computer Vision & Pattern Recognition*, vol. 21, no. 5, pp. 1154–1160, 2011.

[15] T. Mallick, P. P. Das, and A. K. Majumdar, "Characterizations of noise in kinect depth images: A review," *Sensors Journal IEEE*, vol. 14, no. 6, pp. 1731–1740, 2014.

[16] Altmann, Maccarone, McCarthy, Aongus, Newstadt, Gregory, Buller, S. Gerald, and McLaughlin, "Robust spectral unmixing of sparse multispectral lidar waveforms using gamma markov random fields," *IEEE Transactions on Computational Imaging*, vol. PP, no. 99, pp. 1–1, 2016.

[17] Z. Xiao, H. Li, D. Zhou, Y. Dai, and B. Dai, "Accurate extrinsic calibration between monocular camera and sparse 3d lidar points without markers," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017.

[18] J. Sock, J. Kim, J. Min, and K. Kwak, "Probabilistic traversability map generation using 3d-lidar and camera," in *IEEE International Conference on Robotics & Automation*, 2016.

[19] L. Xiao, B. Dai, L. D, T. Hu, and T. Wu, "Crf based road detection with multi-sensor fusion," in *Intelligent Vehicles Symposium*, 2015, pp. 192–198.

[20] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. R. Bradski, "Self-supervised monocular road detection in desert terrain," in *Robotics: Science & Systems Ii, August, University of Pennsylvania, Philadelphia, Pennsylvania, Usa*, 2006.

[21] Q. Zhang, L. Yong, Y. Liao, and W. Yue, "Traversable region detection with a learning framework," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015, pp. 1678–1683, 2015.

[22] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[23] H. Wang, G. Yan, L. Yong, and M. Ren, "Road detection via superpixels and interactive image segmentation," in *IEEE International Conference on Cyber Technology in Automation*, 2014.

[24] El-Naqa, Issam, Yongyi, Yang, Wernick, Miles, N., Galatsanos, Nikolas, and P., "A support vector machine approach for detection of microcalcifications." *IEEE Transactions on Medical Imaging*, 2002.

[25] M. Suzuki, E. Terada, T. Saitoh, and Y. Kuroda, "Vision based far-range perception and traversability analysis using predictive probability of terrain classification," in *ISR/ROBOTIK 2010, Proceedings for the joint conference of ISR 2010 (41st Internationel Symposium on Robotics) und ROBOTIK 2010 (6th German Conference on Robotics), 7-9 June 2010, Munich, Germany - Parallel to AUTOMATICA*, 2011.