2016 IEEE Intelligent Vehicles Symposium (IV)
Gothenburg, Sweden, June 19-22, 2016

Cross Datasets Vegetation Detection with Spatial Prior and Local Context

Heng Fan[1], Xue Mei[2], Danil Prokhorov[2] and Haibin Ling[1]
[1]Computer and Information Sciences Department, Temple University, Philadelphia, PA USA
[2]Toyota Research Institute, North America, Ann Arbor, USA
Email: {hengfan, hbling}@temple.edu, {xue.mei, danil.prokhorov}@toyota.com

Abstract— **In this paper, we propose a vision-based approach for roadside vegetation detection by superpixel matching with local context. Unlike previous detection methods which seek help from additional sensors such as lidar, our algorithm only requires an off-the-shelf camera. The proposed method contains two stages. In the first stage, a superpixel database is constructed by segmenting training images into superpixels, and each superpixel patch is represented with multiple features. After that, the appearance information of vegetation or non-vegetation is encoded in the superpixel database. In the second stage, vegetation detection in each testing image is achieved by superpixel matching. The test image is segmented into superpixels and the (vegetation) label cost of each superpixel is derived by comparing with the k-nearest neighbors in the superpixel database. Furthermore, we incorporate the local context information through the feedback to refine superpixel matching. Taking this context information into account, Markov Random Field (MRF) is utilized to further improve the classification accuracy. Besides, considering the stable layout of road scene images, we utilize spatial priors of road scene to guide vegetation classification. Experiments on real-world datasets demonstrate the promise of our method.**

I. Introduction

For safe navigation the autonomous vehicle may need to be able to detect vegetation on the side of the road as it travels along (see Figure 1). Vegetation detection also helps to match the current environment with the navigation map for localization [12]. To address the problem of vegetation detection, numerous methods have been proposed [2], [5], [12]–[15], [19], [21].

One of the remarkable methods of vegetation detection for outdoor navigation is proposed by Bradley et al. [5]. The method utilizes multiple spectrum camera to generate thermal images for analysis and vegetation detection. Despite generating promising results, the method is unstable in the presence of illumination variations because it depends on trusted data acquisition. In [21], Wellington et al. propose a generative model to detect vegetation which exploits natural structure inherent in outdoor domains. Vandapel et al. [19] suggest a vegetation detection method by interpreting laser data. This approach utilizes 3D points statistics to compute saliency features that capture the spatial distribution of points in a local neighborhood, which is computationally expensive. In [12], Nguyen et al. propose a 2D-3D combination algorithm which is able to utilize complement of three-dimensional point distribution and color descriptor. In [13], a spreading algorithm is suggested for vegetation detection in a cluttered outdoor environment. In [15], Nguyen et al. present a detection approach which enables a double-check process for vegetation detection done by a multi-spectral method while focusing on passable vegetation detection.

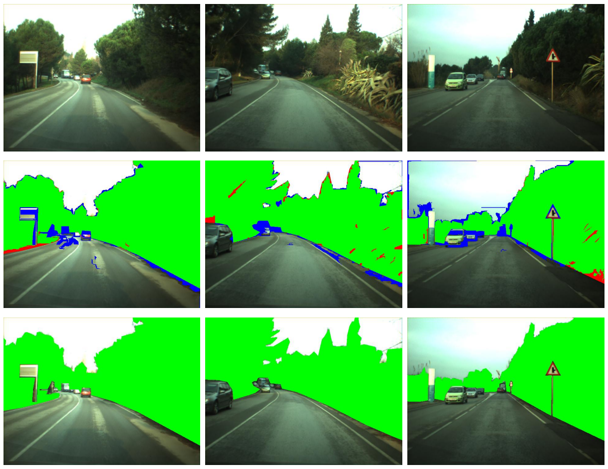

Fig. 1. Detection results on testing images illustrating performance of our approach. First row: testing images. Second row: detection results. Third row: ground truth. In detection results, the green region is vegetation area which is correctly detected, blue region represents non-vegetation area which is mistakenly detected as vegetation, and red region is vegetation area which is not detected. It can be observed that some objects (e.g., vehicles and roadside signs) are misclassified as vegetation, representing areas for future improvements.

Despite promising results for vegetation detection, the aforementioned approaches seek help of non-vision sensors such as lasers. In this work, by contrast, our goal is to try solving the problem of roadside vegetation detection purely by computer vision, requiring nothing more than an ordinary camera.

We propose a novel vision-based vegetation detection method. The problem of vegetation detection is formulated as a classification task which differentiates the vegetation pixels from non-vegetation pixels. This classification process consists of two stages. In the first stage, we construct a superpixel dataset by assembling the superpixels from training images. For robust matching, each superpixel is represented with multiple features. In this way, the appearance information of vegetation and non-vegetation is encoded in the superpixel database. In the second stage, a test image is firstly

978-1-5090-1821-5/16/$31.00 ©2016 IEEE

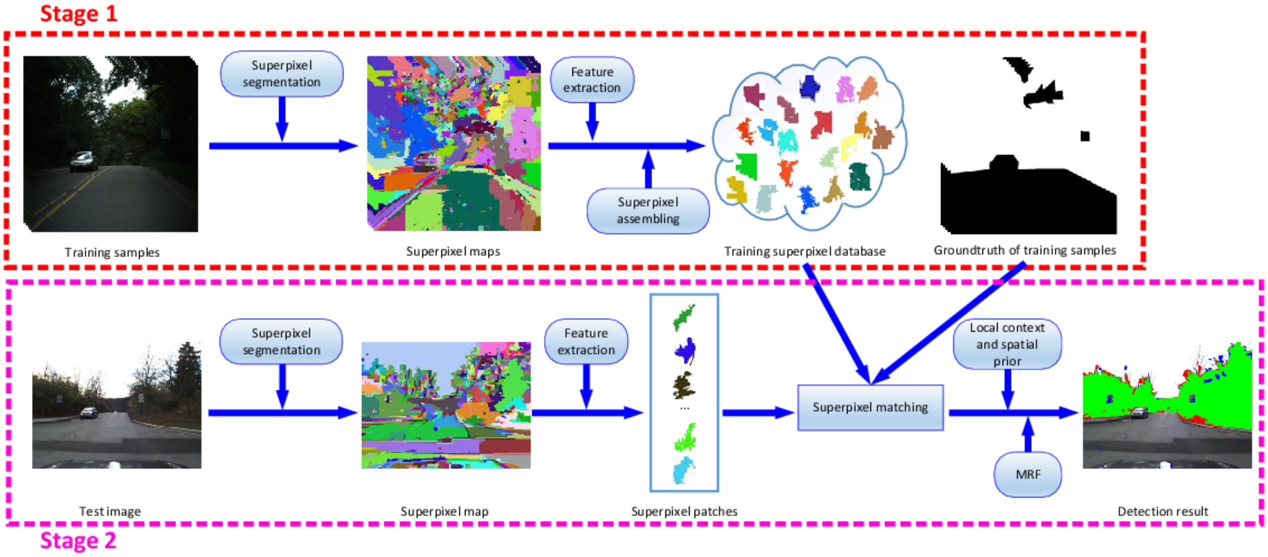

Fig. 2. Framework of the proposed method. In detection result, the green region is vegetation area which is correctly detected, blue region represents non-vegetation area which is mistakenly detected as vegetation, and red region is vegetation area which is not detected.

segmented into superpixels. Then, the vegetation label cost of each superpixel is derived from its k-nearest neighbors in the superpixel dataset built in the first stage. Each pixel inherits the label cost from the superpixel it belongs to. Furthermore, to refine superpixel matching, we incorporate the local context information through a feedback mechanism. Besides, considering the stable layout of road scene images, we exploit spatial prior of road scene to guide vegetation classification. Finally, all these information is unified in the MRF inference framework to achieve pixel classification. Figure 2 illustrates the framework of our method.

Our method does not require additional information from non-vision sensors. For this reason it is difficult to benchmark it against other methods. In our experiments, we use three cross-validation datasets which consist of three different real-world datasets. Experimental results illustrate the effectiveness of the proposed algorithm for vegetation detection.

Our contributions can be summarized as follows:

- A framework for vegetation detection based only on visual inputs is proposed and tested in a cross-validation setting.
- Taking the benefit of relative stability of road scene layout, we utlize spatial prior of road scene to guide vegetation classification.
- A feedback mechanism is adopted to refine the superpixel matching through incorporating the local context information, thereby resulting in further improvement of final classification results.

The rest of this paper is organized as follows. In Section II, we introduce the proposed vegetation detection algorithm in details. Section III presents experimental results and evaluation of the proposed methods, followed by conclusion in Section IV.

II. The Proposed Vegetation Detection Algorithm

In this section, we will introduce our detection approach in details. Section II-A describes the process of constructing training superpixel database and superpixel matching. Incorporation of local context information is introduced in Section II-B. Section II-C exploits the spatial prior of road scene images to guide vegetation detection. And, all these components are integrated in the MRF framework introduced in Section II-D.

A. Superpixel dataset construction and superpixel matching

To develop the superpixel dataset, we oversegment the training images to generate N superpixels using the algorithm in [9]. For each superpixel s_i, we extract four kinds of features including the SIFT histogram [11], RGB histogram, location histogram and PHOG histogram [3]. These histograms are concatenated to represent a superpixel similar as in [23]. The SIFT descriptors of four scales per four pixels are extracted using VLFeat[1] and encoded with five words from a vocabulary of size 1024 by the LLC algorithm [20]. Let x_i denote the feature of s_i, and y_i represents its label, where $y_i \in \{0,1\}$ (1 and 0 represent the vegetation and non-vegetation labels respectively) and is determined by

$$y_i = \begin{cases} 1, & a_i \geqslant 95\% \\ 0, & otherwise \end{cases} \tag{1}$$

where a_i represents vegetation area ratio of s_i. We collect all training superpixels into a database $D = \{s_i, x_i, y_i\}_{i=1}^{N}$.

For each test image, let M be the number of its superpixels. For superpixel s_j (1$\leq$ j $\leq$ M), we compute its

[1] VLFeat is an open source library and available at http://www.vlfeat.org/.

(vegetation) *label cost* by its k-nearest neighbors $\mathcal{N}_k(j)$ (k is set to 7) in D as the following

$$U(y_j = c|s_j) = 1 - \frac{\sum_{i \in \mathcal{N}_k(j), y_i = c} \mathcal{K}(x_j, x_i)}{\sum_{i \in \mathcal{N}_k(j)} \mathcal{K}(x_j, x_i)} \tag{2}$$

where x_j denotes the feature of s_j, $c \in \{0, 1\}$ represents the label and $\mathcal{K}(x_j, x_i)$ is the intersection kernel[2] between x_j and x_i.

B. Local context descriptor

Context is a crucial feature for image classification. We adopt a simple yet effective feedback mechanism as in [17], [18], [23] to account for this information. In the feedback process, we are able to obtain the pixel-wise classification likelihood of each pixel with

$$\ell(p, c) = \frac{1}{1 + \exp(U(y_p = c))} \tag{3}$$

where $U(y_p = c)$ is the cost of assigning label c to pixel p in Eq (9).

For robust superpixel matching, we exploit the local context of each superpixel. For superpixel s_i, we divide its neighborhood into left, right, top, bottom cells $\{lc_i^1, lc_i^2, lc_i^3, lc_i^4\}$. For each cell lc_i^k ($1 \leq k \leq 4$), we compute its sparse context $h_i^k = [h_{i1}^k, h_{i2}^k]$ by

$$h_{ic}^k = \max_{p \in lc_i^k} \ell(p, c) \tag{4}$$

where $\ell(p, c)$ represents the pixel-wise classification likelihood obtained by Eq (3). For superpixel s_i, we can obtain spatial context descriptor $h_i = [h_i^1; h_i^2; h_i^3; h_i^4]$. Thus we can classify the superpixels of test image by Eq (2) with new feature $[x_i; h_i]$.

C. Spatial prior

For road scene images captured by vehicle mounted cameras, it is easy to observe that their scenes share relatively stable layout. Roughly speaking, the upper and bottom parts of a road scene is mostly taken up by non-vegetation areas, while the center part may be a vegetation area. Taking into consideration this stable layout of road scene images, we utilize spatial prior of road scene to guide vegetation classification. We use spatial distribution to encode spatial priors as in [7], [10]. For class c at pixel p, its spatial prior histogram $h_c(p)$ is derived from the training set by

$$h_c(p) = \frac{\sum_{k=1}^{n} f_k^c(p)}{n} \tag{5}$$

where n denotes the number of training samples, and $f_k^c(p)$ is an indicator which indicates whether the label of pixel p in training sample k is c. If its label is c, we set $f_k^c(p)$ to 1, otherwise $f_k^c(p)$ is equal to 0.

The spatial prior term, which indicates the probability that class c is at p is given by

$$E_{sp}(y_p = c) = -\log h_c(p) \tag{6}$$

In MRF inference, this spatial prior information will be incorporated. Figure 3 visualizes the spatial prior of each cross dataset.

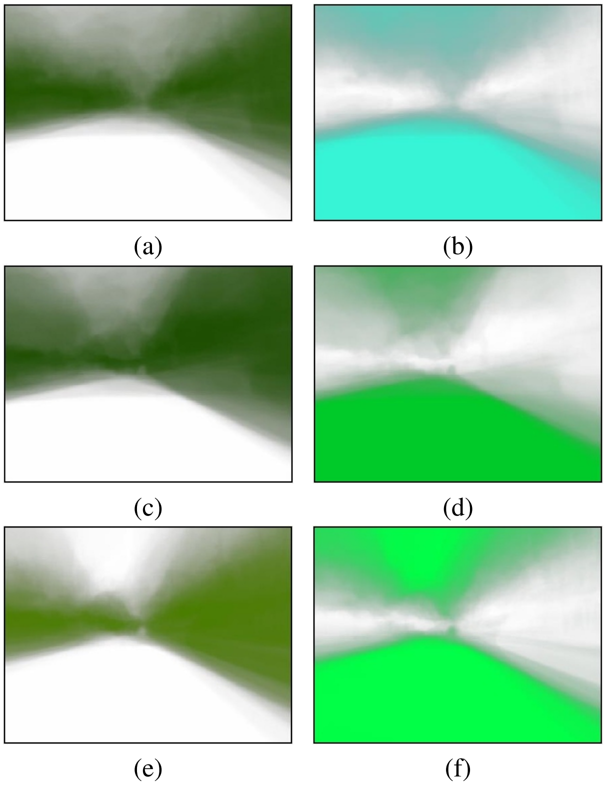

Fig. 3. Spatial priors of vegetation and non-vegetation in three cross datasets. Images (a) and (b) are the spatial priors of vegetation and non-vegetation in CrossDataset1, (c) and (d) in CrossDataset2 and (e) and (f) in CrossDataset3.

D. Markov random field (MRF)

To exploit the context relationship between vegetation pixels and non-vegetation pixels, MRF inference is utilized for contextual constraints. The energy function is given by

$$E(Y) = \sum_p U(y_p = c) + \lambda \sum_{pq} V(y_p = c, y_q = c') \tag{7}$$

where p, q are pixel indices, c, c' are candidate labels and λ is the weight of pairwise energy. Considering the spatial prior of road scene, we revise Eq (7) by incorporating spatial prior term as follows

$$\begin{aligned} E(Y) = & \sum_p U(y_p = c) + \lambda \sum_{pq} V(y_p = c, y_q = c') \\ & + \alpha \sum_p E_{sp}(y_p = c) \end{aligned} \tag{8}$$

where α is a weight (α is set to 0.1). The unary energy of one pixel is given by its superpixel

$$U(y_p = c) = U(y_j = c|s_j), \; p \in s_j \tag{9}$$

The pairwise energy on edges is given by spatially variant label cost

$$V(c, c') = d(p, q) \cdot \mu(c, c') \tag{10}$$

[2] $\mathcal{K}(x_j, x_i)$ is defined as $\mathcal{K}(x_j, x_i) \approx < \phi(x_j), \phi(x_i) >$ in VLFeat.

where $d(p,q) = exp(-\|I(p)-I(q)\|^2/2\sigma^2)$ (σ is the standard deviation) is the color dissimilarity between two adjacent pixels, and $\mu(c,c')$ is the penalty of assigning label c and c' to two adjacent pixels and defined by log-likelihood of label co-occurrence statistics

$$\mu(c,c') = -\log[(P(c|c') + P(c'|c))/2] \times \sigma \tag{11}$$

In this way, we are able to obtain the labels of all pixels by performing MAP inference on $E(Y)$ by graph cut optimization in [4].

Finally, we summarize the proposed vegetation detection system in Algorithm 1.

Algorithm 1 The proposed algorithm

```
Stage 1: Construction of superpixel database
    Segment training images into superpixels using algorithm in [9];
    Extract multiple features for each superpixel $s_i$;
    Concatenate these features into a combined feature $x_i$ to represent $s_i$, and $y_i$ is its label;
    Collect all the training superpixels to construct the superpixel database $D = \{s_i, x_i, y_i\}_{i=1}^N$;
Stage 2: Vegetation detection for a testing image
    Segment the testing image into $M$ superpixels by algorithm in [9];
    Compute the feature $x_j$ for superpixel $s_j$;
    Compute the label cost of $s_j$ by its k-nearest neighbors $\mathcal{N}_k(j)$ in $D$ based on Eq (2);
    Compute the pixel-wise classification likelihood of each pixel based on Eq (3);
    Compute local context based on Eq (4);
    Calculate the spatial prior histogram $h_c(p)$ of the training set based on Eq (5);
    Compute the spatial prior term based on Eq (6);
    Perform MRF based on the revised energy function $E(Y)$ in Eq (8);
    Obtain the classification result, and segment vegetation area from non-vegetation region;
```

III. Experimental Results and Analysis

A. Dataset

We use three datasets denoted as KITTI-Veggie[3], TOYOTA-Veggie1 and TOYOTA-Veggie2[4] to generate three cross datasets which are CrossDataset1 (CD1), CrossDataset2 (CD2) and CrossDataset3 (CD3) to test the proposed algorithm. Table I describes the attributes of each dataset, and Figure 4 shows some samples from each dataset.

To generate cross datasets, we treat two of the three datasets as training samples and the rest as testing samples. We are able to obtain three cross datasets, i.e., CrossDataset1, CrossDataset2 and CrossDataset3. For CrossDataset1, TOYOTA-Veggie1 and TOYOTA-Veggie2 are training samples, and KITTI-Veggie is testing samples. For CrossDataset2, TOYOTA-Veggie1 and KITTI-Veggie are training samples, and TOYOTA-Veggie2 is testing samples. For CrossDataset3, TOYOTA-Veggie2 and KITTI-Veggie are training samples, and TOYOTA-Veggie1 is testing samples.

TABLE I
Description of the three datasets.

Name	Frames	RGB	Size
KITTI-Veggie	303	Yes	640 × 480
TOYOTA-Veggie1	434	Yes	512 × 512
TOYOTA-Veggie2	346	Yes	484 × 364

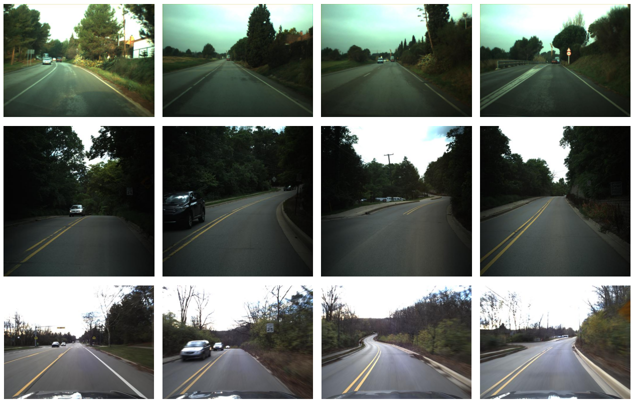

Fig. 4. Image samples of each dataset. First row: image samples in KITTI-Veggie. Second row: image samples in TOYOTA-Veggie1. Third row: image samples in TOYOTA-Veggie2.

B. Experimental results

The proposed algorithm is implemented in MATLAB on a 3.2 GHz Intel E3-1225 v1 Core PC with 8GB memory. We run the proposed method on the three cross datasets respectively. Table II shows the confusion matrices [13] of the proposed method on the three cross datasets.

TABLE II
Confusion matrices of the proposed method.

		Ground truth Vegetation (%)	Ground truth Non-Vegetation (%)
CD1[a]	Vegetation	97.3	2.7
	Non-Vegetation	9.66	90.34
CD2[b]	Vegetation	90.14	9.86
	Non-Vegetation	3.56	96.44
CD3[c]	Vegetation	97.66	2.34
	Non-Vegetation	11.48	88.52
Avg	Vegetation	95.03	4.97
	Non-Vegetation	8.23	91.77

[a] CD1: CrossDataset1.
[b] CD2: CrossDataset2.
[c] CD3: CrossDataset3.

For detailed analysis, we report our results in the following

[3] KITTI-Veggie is collected from part of KITTI Vision Benchmark which is available at http://www.cvlibs.net/datasets/kitti/.

[4] TOYOTA-Veggie1 and TOYOTA-Veggie2 are collected by Toyota Technical Center.

formats: the Vegetation ∼ Ground truth Vegetation (V ∼ GV), Vegetation ∼ Ground truth Non-Vegetation (V ∼ GNV), Non-Vegetation ∼ Ground truth Vegetation (NV ∼ GV) and Non-Vegetation ∼ Ground truth Non-Vegetation (NV ∼ GNV) for each image (frame) in the three cross datasets. Figure 5, 6 and 7 demonstrate the V ∼ GV, V ∼ GNV, NV ∼ GV and NV ∼ GNV of each image in CrossDataset1, CrossDataset2 and CrossDataset3. Figure 8, 9 and 10 show the detection results of some testing samples in CrossDataset1, CrossDataset2 and CrossDataset3 with our method.

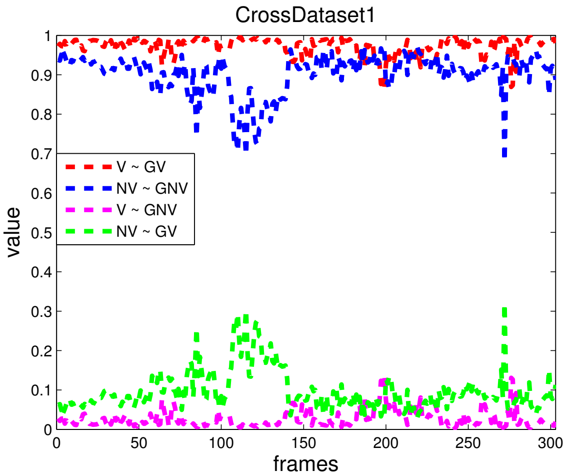

Fig. 5. The V∼ GV, V∼ GNV, NV∼ GV and NV∼ GNV of each image in CrossDataset1. The order of frames (X axis) is the same as that in the original video sequences.

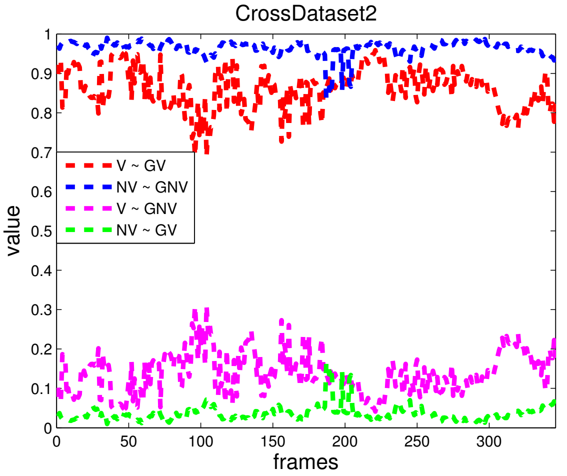

Fig. 6. The V∼ GV, V∼ GNV, NV∼ GV and NV∼ GNV of each image in CrossDataset2. The order of frames (X axis) is the same as that in the original video sequences.

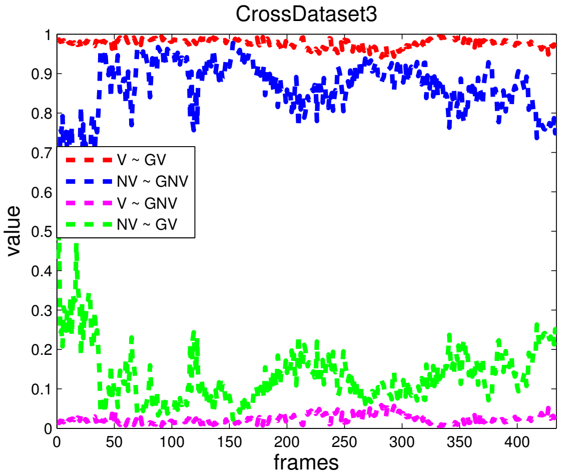

Fig. 7. The V∼ GV, V∼ GNV, NV∼ GV and NV∼ GNV of each image in CrossDataset3. The order of frames (X axis) is the same as that in the original video sequences.

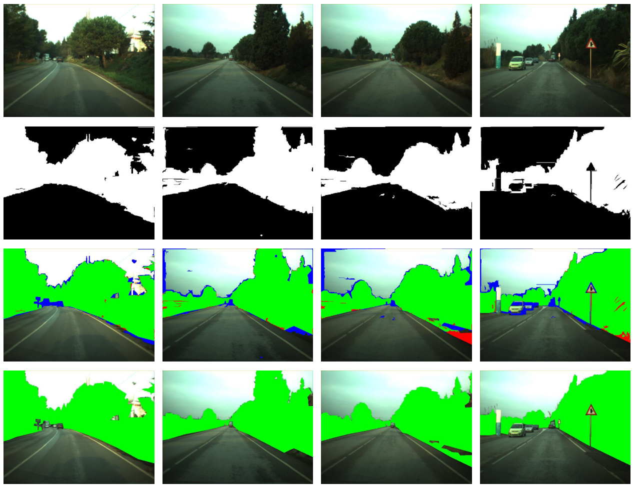

Fig. 8. Detection results of testing images in CrossDataset1. First row: original images. Second row: classification results. Third row: detection results. Fourth row: groundtruth. In detection result, the green region is vegetation area which is correctly detected, blue region represents non-vegetation area which is mistakenly detected as vegetation, and red region is vegetation area which is not detected.

IV. Conclusion

In this paper we propose a novel vegetation detection method using superpixel matching with spatial prior and local context. Unlike previous works, our method utilizes computer vision exclusively, which might provide overall simplification of the perception system. To evaluate performance of the proposed approach, we construct three cross-validation datasets and test the proposed algorithm on them. Experiments demonstrate that the proposed method is promising for roadside vegetation detection. It might be possible to apply the same method to detecting other large roadside structures, e.g., building, in dense urban environment. For future work, the detection of vehicles and traffic signs on the road need to be considered to improve accuracy of the method. Besides, we will conduct more experiments and compare our approach with other vegetation detection methods to demonstrate the effectiveness of the proposed algorithm.

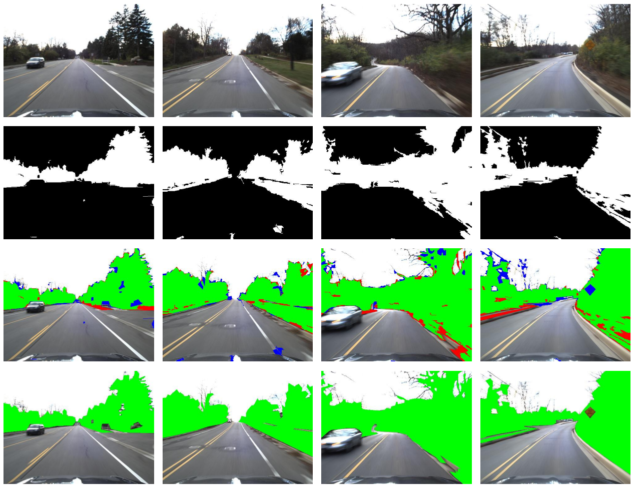

Fig. 9. Detection results of testing images in CrossDataset2; the coloring is the same as in the previous figure.

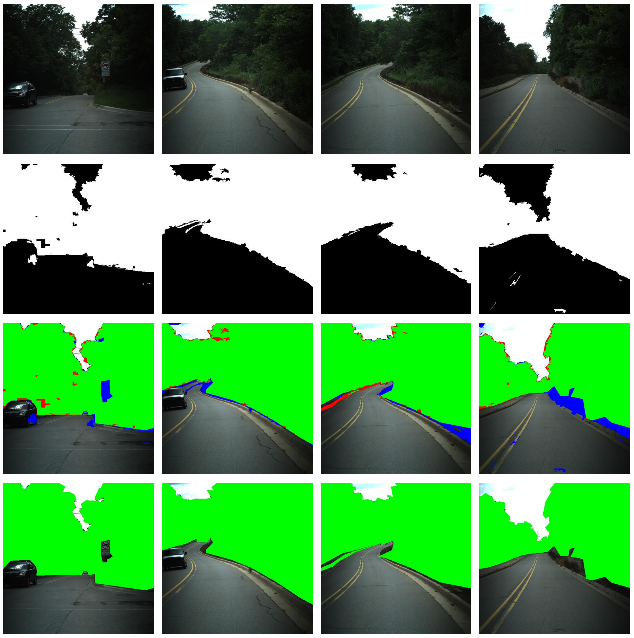

Fig. 10. Detection results of testing images in CrossDataset3; the coloring is the same as in the previous figure.

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012.

[2] A. Angelova, L. Matthies, D. Helmick, and P. Perona, "Fast terrain classification using variable-length representation for autonomous navigation," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

[3] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *ACM international conference on Image and video retrieval*, pp. 401-408, 2007.

[4] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222-1239, 2001.

[5] D. M. Bradley, R. Unnikrishnan, and J. Bagnell, "Vegetation detection for driving in complex environments," in *IEEE International Conference on Robotics and Automation*, pp. 503-508, 2007.

[6] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603-619, 2002.

[7] S. Di, H. Zhang, X. Mei, D. Prokhorov and H. Ling. "Spatial prior for nonparametric road scene parsing," in *IEEE Intelligent Conference on Intelligent Transportation Systems*, pp. 1209-1214, 2015.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 886-893, 2005

[9] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167-181, 2004.

[10] C. Lie, J. Yuen, and A. Torralba. "Nonparametric scene parsing via label transfer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2368-2382, 2011.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[12] D. V. Nguyen, L. Kuhnert, T. Jiang, S. Thamke, and K. D. Kuhnert, "Vegetation detection for outdoor automobile guidance," in *IEEE International Conference on Industrial Technology*, pp. 358-364, 2011.

[13] D. V. Nguyen, L. Kuhnert, and K. D. Kuhnert, "Spreading algorithm for efficient vegetation detection in cluttered outdoor environments," *Robotics and Autonomous Systems*, vol. 60, pp. 1498-1507, 2012.

[14] D. V. Nguyen, L. Kuhnert, and K. D. Kuhnert, "Structure overview of vegetation detection. A novel approach for efficient vegetation detection using an active lighting system," *Robotics and Autonomous Systems*, vol. 60, pp. 498-508, 2012.

[15] D. V. Nguyen, L. Kuhnert, S. Thamke, J. Schlemper, and K. D. Kuhnert, "A novel approach for a double-check of passable vegetation detection in autonomous ground vehicles," in *IEEE International Conference on Intelligent Transportation Systems*, pp. 230-236, 2012.

[16] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[17] G. Singh and J. Kosecka, "Nonparametric scene parsing with adaptive feature relevance and semantic context," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3151-3157, 2013.

[18] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3d brain image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1744-1757, 2010.

[19] N. Vandapel, D. F. Huber, A. Kapuria, and M. Hebert, "Natural terrain classification using 3-d ladar data," in *IEEE International Conference on Robotics and Automation*, pp. 5117-5122, 2004.

[20] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality constrained linear coding for image classification," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3360-3367, 2010.

[21] C. Wellington, A. Courville, and A. Stentz, "A generative model of terrain for autonomous navigation in vegetation," *International Journal of Robotics Research*, vol. 25, no. 12, pp. 1287-1304, 2006.

[22] F. Yang, H. Lu, and M.-H. Yang, "Robust superpixel tracking," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1639-1651, 2014.

[23] J. Yang, B. Price, S. Cohen, and M.-H. Yang, "Context driven scene parsing with attention to rare classes," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3294-3301, 2014.