ADAPTIVE ROAD DETECTION TOWARDS MULTISCALE-MULTILEVEL PROBABILISTIC ANALYSIS

Zhiyu Jiang[1,3], *Qi Wang*[2,*], *Yuan Yuan*[1]

[1]Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China.

[2]Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.

[3]Graduate University of the Chinese Academy of Sciences, 19A Yuquanlu, Beijing, 100049, P. R. China.

ABSTRACT

Vision-based road detection is a challenging problem because of the changeable shape and varying illumination. Though many efforts have been spent on this topic, the achieved performance is far from satisfactory. To this end, this paper formulates a Bayesian method which simultaneously explores the multiscale-multilevel clues that are considered to be complementary. Two contributions are claimed in this proposed method. 1) By computing the prior distribution in superpixel-level with a novel *Laplacian Sparse Subspace Clustering* and observation likelihood in pixel-level with statistical color similarity, the posterior probability of road region can be effectively inferred. 2) To ensure the adaptivity of road model in various conditions, a multiscale strategy is presented to fuse the detection results of different scales. Experimental results on several challenging video sequences verify the superiority of the proposed method compared with several popular ones.

Index Terms— Computer vision, road detection, Bayesian, sparse, clustering, superpixel

1. INTRODUCTION

Road detection plays a key role for enabling the *Driver Assistance Systems* (DAS), and can provide a significant contextual clue for target detection (e.g, vehicles and pedestrians) [1, 2, 3, 4, 5, 6]. Although significant progress on road detection has been made, it is still a challenge due to the varying illumination and road appearance change. Towards tackling these difficulties, existing methods can be divided into two groups: *feature-based* and *model-based*.

Feature-based method [3, 7, 8] relies upon the extraction of image local features which are utilized to detect road boundary or road region. The features such as color, gradient and texture are commonly used by feature clustering [9], threshold segmentation[10] or region growing [2] to infer the road area. The main advantages of *feature-based method* are that it is insensitive to the road shape and little previous knowledge is needed. However, different feature selections can make the detection results different and the method is sensitive to shadows and other illumination changes.

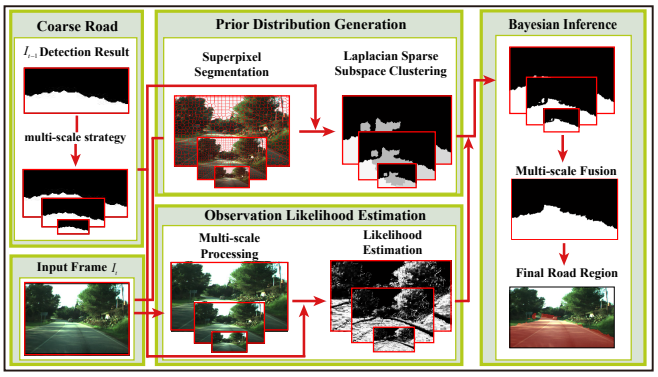

Fig. 1. Road detection pipeline.

Model-based method [11, 12] tends to have an assumption of the road shape based on the prior knowledge of the road shape. Hence the aim of road detection problem is converted into finding the fittest parameters under the model assumption. Although *model-based method* can accurately determine the road region given a proper road model, it may be invalid to face the situation where road shapes change as the vehicle is moving. Therefore, it is difficult to find an appropriate model for unstructured roads with inconstant conditions.

For tackling the above issues, a novel Bayesian inference is utilized in this work to combine the multiscale detection results generated by multilevel (superpixel-level and pixel-level) cues. The proposed method incorporates advantages of model based and feature based methods, and proves to be effective and robust.

This work is supported by the State Key Program of National Natural Science of China (Grant No. 61232010), and the National Natural Science Foundation of China (Grant No. 61172143, 61379094 and 61105012).

1.1. Overview of the Proposed Method

The main steps of road detection method proposed in this paper are illustrated in Fig. 1. By modeling the road detection problem as the posterior probability estimation under the Bayesian framework, the key issues of road detection become the prior distribution generation and the observation likelihood estimation in the image plane.

Before finely delineate the road region, the coarse road region is firstly extracted based on the temporal constraints between adjacent frames. After that, the image is represented by superpixels for the reduction of computation complexity and elimination of noisy pixels. Then the prior probability distribution is estimated by computing the overlapping ratio of the superpixel clustering results and the previously extracted coarse road region. At the same time, the observation likelihood is estimated based on the color histograms of the road/non-road regions separated by the coarse separation. With the generated prior distribution and the estimated observation likelihood, the posterior probability of road can be naturally inferred based on the Bayesian framework. Considering the fact that the road region may not be represented at a proper scale, it is necessary to analyze the multiscale situation. For this reason, the final road region is obtained by linearly fusing the posterior probabilities under different image scales.

1.2. Contributions

The proposed method is novel and effective. We claim two contributions in this procedure.

1) For modeling the road prior, *Laplacian Sparse Subspace Clustering* (LSSC) is for the first time introduced to cluster the superpixel-level feature space. Compared to other methods for road prior modeling [9, 2, 10], LSSC can effectively explore the structure manifold of data and is robust to outliers. As for the observation likelihood, the pixel-level statistics are utilized to embody the details of the image. The fusion of multilevel information in different levels makes the road inference more stable.

2) To detect the road more accurately, this paper formulates a multiscale fusion strategy to refine the detection result. Different from other methods that only emphasize on road region at a single scale [9, 2], we simultaneously take the road and non-road regions at different scales into consideration. This processing make the detected road region be more accurate.

The remainder of this paper is organized as follows. In section 2, the proposed method is described in detail. In section 3, experimental results are presented and evaluated on several challenging video sequences. Finally, conclusion is made in section 4.

2. PROPOSED METHOD

In this section, we present the effective Bayesian road detection method. It is achieved by the prior knowledge generation by superpixel-level feature clustering and observation likelihood estimation via pixel-level color evaluation. To be specific, given the current frame I_t and the coarse road region c_t, the road detection can be formulated as calculating the probability of each pixel x_t belonging to the road r_t

$$p(r_t|x_t) = \frac{p(r_t)p(x_t|r_t)}{p(r_t)p(x_t|r_t)+p(\widetilde{r_t})p(x_t|\widetilde{r_t})}, \quad (1)$$

where $p(r_t|x_t)$ represents the posterior probability for predicting a pixel as road region, $p(r_t)$ and $p(\widetilde{r_t}) = 1 - p(r_t)$ specify the prior knowledge of road/non-road area, $p(x_t|r_t)$ and $p(x_t|\widetilde{r_t})$ indicate the observation likelihood of the road/no-road. With this definition, the final decision is made with a threshold 0.5. Larger posterior than this value is claimed to be the road area. Therefore, the key issue of the proposed method is how to generate the prior $p(r_t)$ and the observation likelihoods $p(x_t|r_t)$ and $p(x_t|\widetilde{r_t})$.

2.1. Coarse Road Region Extraction

With the temporal constraints between frames, the coarse road region in current frame is initialized based on the previously detected result. In other words, the current road estimation can be regarded as the deformation of road region at previous frame. As for the first frame, a rectangular region of certain size in the middle and low part of the image plane is defined as the latent road, which is always the case for plenty of road videos. Then this region is treated as the seed and expanded according to the color similarity. The obtained result is taken as the coarse road region.

2.2. Prior Knowledge Generation by LSSC

With the obtained coarse road region, the next step is to get the prior knowledge. For this purpose, superpixels are first acquired from the input frames and then clustered by LSSC. The final road prior can be inferred by combining the clustering result and the coarse road region.

2.2.1. Superpixel clustering via LSSC

In order to avoid the influence of the noisy pixels and reduce the computing complexity, the examined frame I_t is segmented into N superpixels by the robust TurboPixels method [13]. The feature vector of the i^{th} superpixel sp_i is represented as $\mathbf{u}_i, i = 1, ..., N$, which is constructed by the average value of the intensity, color and gradient of the pixels within sp_i, and $\mathbf{u}_i \in \mathbb{R}^{D\times 1}$ is specified as

$$\mathbf{u}_i = [\mathbf{C}_i; \mathbf{I}_i; \mathbf{G}_i], \quad (2)$$

where D is the dimension of the feature vector, and $D \ll N$, $\mathbf{C}_i, \mathbf{I}_i$, and $\mathbf{G}_i$ represent the color, intensity and gradient information of the superpixel.

The representation of superpixel can eliminate the influence of noise greatly. However, the illumination can still bring difficulties for the processing. To this end, *Laplacian Sparse Subspace Clustering* (LSSC) [14] algorithm is utilized to group superpixels considering its robustness and effectiveness to varying illumination conditions [15]. The algorithm is motivated by the assumption that each data point belongs to a unique subspace and it can be represented by a small set of points of the subspace. That means each point has a sparse representation considering the entire set of data points. By searching for the sparsest combination of each point, the subspace that a certain point belongs to would be determined [15] accordingly.

According to the theory mentioned above, the sparse representation of the superpixel $\mathbf{u}_i$ can be inferred by the following formula

$$\min \|\mathbf{c}_i\|_1 \quad s.t. \quad \mathbf{u}_i = \mathbf{U}_i \mathbf{c}_i \ \ and \ \ \mathbf{c}_i^\top \mathbf{1} = 1, \tag{3}$$

where $\mathbf{U}_i = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\} \in \mathbb{R}^{D \times (N-1)}$ is the data array of $N-1$ points by removing $\mathbf{u}_i$. Generally, $\mathbf{u}_i$ is corrupted by a small disruption η_i. Therefore, the observation vector is modeled by $\mathbf{y}_i = \mathbf{u}_i + \eta_i$. Solving $\mathbf{c}_i$ is equivalent to minimize the following objective function

$$\min \|\mathbf{U}_i \mathbf{c}_i - \mathbf{y}_i\|_2 + \lambda \|\mathbf{c}_i\|_1 \quad s.t. \quad \mathbf{c}_i^\top \mathbf{1} = 1. \tag{4}$$

Because $D \ll N$, $\mathbf{U}_i$ is an over-complete dictionary for superpixel $\mathbf{u}_i$ and a small variation of two similar points would lead to a drastic difference to the basis. Inspired by that, this paper introduces a Laplacian regularization term [14] to enforce that similar superpixels should have similar sparse coefficients. Therefore, Eq. 4 is rewritten as

$$\begin{aligned} &\min \|\mathbf{U}_i \mathbf{c}_i - \mathbf{u}_i\|_2 + \lambda \|\mathbf{c}_i\|_1 + \alpha/2 \sum_{ij} \|\mathbf{c}_i - \mathbf{c}_j\|^2 W_{ij} \\ &\quad = \min \|\mathbf{U}_i \mathbf{c}_i - \mathbf{u}_i\|_2 + \lambda \|\mathbf{c}_i\|_1 + \alpha \ tr(\mathbf{C}\mathbf{L}\mathbf{C}^\top) \\ &s.t. \ \ \mathbf{c}_i^\top \mathbf{1} = 1. \end{aligned} \tag{5}$$

where W_{ij} is the constraint matrix denoting the similarity between superpixels, $\mathbf{L}$ is the Laplacian matrix defined as $\mathbf{L} = \mathbf{H} - \mathbf{W}$. Additionaly, $\mathbf{H}$ is the diagonal matrix and $H_{ij} = \sum_j W_{ij}$, and α is the paramenter for balancing the reconstruction error and regularization (set to 0.2 in our experiments). The LSSC can be iteratively solved by [16].

2.2.2. Prior knowledge generation

The LSSC method can properly cluster the image into several partitions which are robust to varying illumination. However, the clustering procedure is unsupervised and the road region is also unknown. Therefore, the previously obtained coarse road is ultilized to define the prior probability

$$p(r_t) = \frac{|partition \cap coarse|}{|partition|}, \tag{6}$$

where $partition$ represents the examined cluster of superpixels, $coarse$ denotes the coarse road region, and $|\,\mathbf{.}\,|$ is the size of the corresponding region. After this process, each cluster (superpixel/pixel) will get a calculated prior.

Table 1. The Data Set Description.

Video	No.	Attributes
srd	4000	T-intersection road
srn	977	Low intensity, street lamp lighting reflection
hr	1238	Large pitch and yaw change, wide baseline
mr	580	No clear boundary, variational road boundary
sr	346	Sunny, shadow
rr	244	Low intensity, reflection

2.3. Observation Likelihood Estimation

The observartion likelihood of each pixel can be estimated by considering the statistical low level features of road and non-road regions simultaneously. The detailed procedure is presented as follows.

According to the obtained coarse road region, the image can be divided into road and non-road regions. In order to judge whether a pixel belongs to the road region or not, the statistic information of color space is utilized to establish the criteria. Let N_1 and N_2 respectively represent the pixel number of coarse road and non-road regions. These regions are represented by a histogram of color space. Given a pixel x_t in frame I_t, $N_{1f(x_t)}, f \in \{r, g, b\}$ denotes the number of road region pixels whose values in color channel f are $f(x_t)$, and $N_{2f(x_t)}$ is similarly defined in the non-road region. For general conditions, the different color channels are assumed to be independent. Therefore, the observation likelihood of pixle x_t is written as

$$\begin{aligned} p(x_t|r_t) &= \prod_f \frac{N_{1f(x_t)}}{N_1}, \\ p(x_t|\widetilde{r_t}) &= \prod_f \frac{N_{2f(x_t)}}{N_2}. \end{aligned} \tag{7}$$

Substituting Eq. 6 and Eq. 7 into Eq. 1, the posterior probability for predicting a superpixel being road region is obtained. Since it is hard to judge whether the road region is properly represented at a certain image scale, the multiscale strategy is adopted. That means the above steps are repeated at each scale and the final result is obtained by thresholding the linear fusion of the posterior probabilities under different image scales.

3. EXPERIMENTS

3.1. Dataset

To prove the effectiveness of the proposed method, two publicly available video sequences [7] named Sunny Road (*sr*)

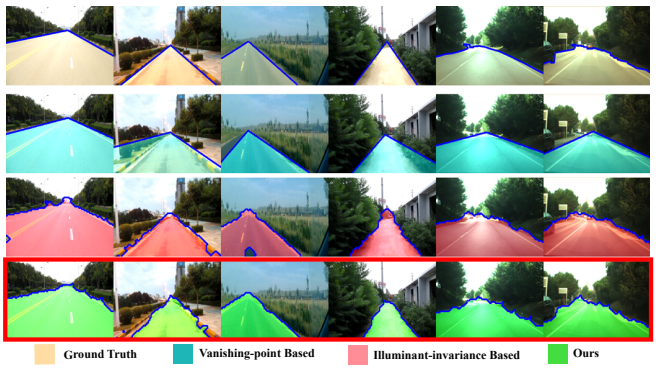

Fig. 2. Typical road detection results of different methods.

and Rain Road (*rr*) are utilized. Besides, four additional videos collected by ourselves from urban and rural areas are also utilized to test the ability of tackling varying illumination, dramatic boundary change and obstacle objects. They are Structural Road in Daytime (*srd*), Structural Road in Nightfall (*srn*), Highway Road (*hr*), and Muddy Road (*mr*). The detailed information of the videos, including the frame number and attribute, is presented in Table 1. In our experiments, all video frames are normalized into 300×500 and the ground truth is labeled every two frames for each video.

3.2. Implementation Setups

Competitors. Two state-of-the-art road detection methods are selected as the competitors. They are Vanishing-point (VP) based method [17] and Illuminant-invariance based (ILL) method [7]. Among them, VP assumes that the road boundary can be fitted by two straight lines and ILL claims to be illumination-invariant. The two methods are employed to prove the ability of the proposed method for handling the changeable road scene and varying illumination of road surface.

Evaluation criterions. For the evaluation of the detection results, we mainly focus on the pixel-wise metrics. *Accuracy*, *precision*, *recall* and *F-measure* are used in this work, which measure the different aspects between the detected road region and the manually labeled ground truth. The four evaluation criterions are defined based on the confusion matrix as described in [18].

3.3. Performance Analysis

In this part, experimental results are evaluated by qualitative and quantitative analysis. Typical road detection results on six sceneries are shown in Fig. 2. From Fig. 2, we can find that the VP method can perform well when the road can be fitted by two straight lines. But the road boundaries are still wrongly estimated in many cases, such as the second and third images. This is because the VP method is based on single image and the temporal constraint is not considered. However, our method can properly handle this situation by utilizing the the coarse road region which is generated from previous frame.

Table 2. Quantitative comparison of different methods. The **bold** figures represent the best scores.

Method	Video	Accuracy	Precision	Recall	F-measure
VP [17]	*srd*	92.05	93.43	90.89	92.12
	srn	82.45	**94.59**	63.41	75.74
	hr	86.81	91.05	67.24	74.67
	mr	85.13	87.90	64.29	73.75
	sr	90.45	91.38	87.02	89.09
	rr	91.10	91.76	88.28	89.84
	average	88.00	91.69	76.86	82.54
ILL [7]	*srd*	92.34	94.88	90.09	92.42
	srn	90.66	94.35	81.07	87.22
	hr	92.16	93.44	81.04	86.72
	mr	92.05	92.47	83.60	87.77
	sr	90.77	94.64	85.15	89.65
	rr	91.19	93.96	86.61	90.13
	average	91.53	93.96	84.59	88.99
Ours	*srd*	**97.24**	**99.30**	**95.49**	**97.34**
	srn	**94.67**	92.61	**87.88**	**89.92**
	hr	**96.87**	**99.59**	**83.20**	**90.49**
	mr	**96.62**	**88.65**	**94.16**	**91.08**
	sr	**95.09**	**95.92**	**91.56**	**93.59**
	rr	**95.28**	**95.40**	**92.56**	**93.82**
	average	**95.96**	**95.25**	**90.81**	**92.71**

The ILL method can have a good performance in the illuminant invariance situation, such as the last two videos. Unfortunately, some road regions is not successfully detected or are falsely detected as non-road regions. However, the proposed method considers the spatial correlation in two aspects, one of which is the superpixel representation of the original image, the other is the the superpixel clustering based on LSSC. The two aspects can promise that the estimated road region is integral for its spatial constraints consideration. The results in Tab. 2 also verify the effectiveness and robustness of the proposed method, from which we can see that the highest scores are mostly achieved by the proposed method and the average performances of the proposed method is boosted between $5\% - 15\%$ on the F-measure criterion.

4. CONCLUSION

In this paper, aiming to handle the problems of changeable shape and varying illumination in road detection, a Bayesian based method is proposed to properly explore the multiscale-multilevel feature analysis. In this procedure, LSSC is first introduced to cluster the superpixels to generate the prior knowledge, which can efficiently discard the outliers and explore the structure manifold of the data. The observation likelihood is simultaneously estimated based on the statistic color similarity both in road and non-road regions. In the end, a multiscale fusion strategy is adopted to further refine the detected road region. Experiments on six typical videos verify the robustness and efficiency of the proposed method.

In this work, our focus is mainly on the road detection with changeable shape and varying illumination. Future work will extent to the situation with more complex traffic environment, such as road detection with vehicles and pedestrians on the road surface.

5. REFERENCES

[1] H. Heng and H. Xiong, "Pedestrian detection based on road surface extraction in pedestrian protection system," in *Mechatronics and Automatic Control Systems*, vol. 237, pp. 793–800. 2014.

[2] F. Oniga, A. Nedevschi, M. Meinecke, and T. Thanh-Binh, "Road surface and obstacle detection based on elevation maps from dense stereo," in *Proc. IEEE Intelligent Transportation Systems Conference*, 2007, pp. 859–865.

[3] M. Sotelo, F. Rodriguez, and L. Magdalena, "Virtuous: Vision-based road transportation for unmanned operation on urban-like scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 2, pp. 69–83, 2004.

[4] Q. Wang, J. Fang, and Y. Yuan, "Multi-cue based tracking," *Neurocomputing*, vol. 131, pp. 227–236, 2013.

[5] Q. Wang, P. Yan, Y. Yuan, and X. Li, "Multi-spectral saliency detection," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 34–41, 2013.

[6] Q. Wang, Y. Yuan, and P. Yan, "Visual saliency by selective contrast," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 7, pp. 1150–1155, 2013.

[7] J. Alvarez and A. Lopez, "Road detection based on illuminant invariance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2011.

[8] N. Salim, X. Cheng, and X. Degui, "A robust approach for road detection with shadow detection removal technique," *Information Technology Journal*, vol. 13, no. 4, pp. 782–788, 2014.

[9] R. Maurya, P. Gupta, and A. Shukla, "Road extraction using k-means clustering and morphological operations," in *Proc. Int' Conf. Image Information Processing*, 2011, pp. 1–6.

[10] C. Rotaru, T. Graf, and J. Zhang, "Color image segmentation in hsi space for automotive applications," *Journal of Real-Time Image Processing*, vol. 3, no. 4, pp. 311–322, 2008.

[11] A. Borkar, M. Hayes, and M. Smith, "Robust lane detection and tracking with RANSAC and kalman filter," in *Proc. IEEE Int' Conf. Image Processing*, 2009, pp. 3261–3264.

[12] H. Sawano and M. Okada, "Road extraction by snake with inertia and differential features," in *Proc. Int' Conf. Pattern Recognition*, 2004, vol. 4, pp. 380–383.

[13] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, and S. Dickinson, "Turbopixles:fast superpixels using geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2290–2297, 2009.

[14] Y. Xie, H. Lu, and M. Yang, "Bayesian saliency via low and mid level cues," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1689–1698, 2013.

[15] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman, "Clustering appearences of objects under varying illumination conditions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003, vol. 1, pp. 11–18.

[16] S. Gao, I. Tsang, L. Chia, and P. Zhao, "Local features are not lonely-laplacian sparse coding for image classification," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010, vol. 1, pp. 3555–3561.

[17] H. Kong, J. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, vol. 1, pp. 96–103.

[18] Q. Wang, Y. Yuan, P. Yan, and X. Li, "Saliency detection by multiple-instance learning," *IEEE Transactions on Cybernetics*, vol. 43, no. 2, pp. 660–672, 2013.