A Statistical Modeling Method for Road Recognition in Traffic Video Analytics

Hang Shi, Hadi Ghahremannezhadand, and Chengjun Liu
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ, 07102 USA
hs328@njit.edu, hg255@njit.edu, cliu@njit.edu

***Abstract*—A novel statistical modeling method is presented to solve the automated road recognition problem for the region of interest (RoI) detection in traffic video cognition. First, a temporal feature guided statistical modeling method is proposed for road modeling. Specifically, a foreground detection method is applied to extract the temporal features from the video and then to estimate a background image. Furthermore, the temporal features guide the statistical modeling method to select sample data. Additionally, a model pruning strategy is applied to estimate the road model. Second, a new road region detection method is presented to detect the road regions in the video. The method applies discrimination functions to classify each pixel in the estimated background image into a road class or a non-road class, respectively. The proposed method provides an intra-cognitive communication mode between the ROI selection and video analysis systems. Experimental results using real traffic videos from the New Jersey Department of Transportation (NJDOT) show that the proposed method is able to (i) detect the road region accurately and robustly and (ii) improve upon the state-of-the-art road recognition methods.**

I. Introduction

Region of interest (RoI) is a widely used concept in video analysis. A region of interest is defined by a subspace of the entire video frame, which includes the part that people are most concerned about. To better recognize the activities in traffic, people always want to apply the video analysis algorithms within the RoI instead of the whole frame to reduce the computation complexity of the video analysis tasks [1]. Therefore, inter-cognitive communication [2], [3] between the human and artificial video analysis systems exists during the ROI selection. With the development of artificial intelligence and cognitive info-communications, an intra-cognitive communication mode [2], [3] becomes a more popular way in ROI detection. The ROI detection system recognizes the ROI and transfer the ROI information to the video analysis system. The communication between the two artificial cognitive systems can largely reduce the unbalanced cognitive capabilities between human beings and video analysis systems.

In traffic surveillance videos, the most widely used RoI is the region of the road. Because most of the traffic activities are happened in the road area, such as traffic congestion, wrong way vehicles, and traffic accidents. Manually selecting the RoI is a common approach when analyzing the traffic surveillance videos. However, people need to select the road region as the RoI for every different camera location, and redefine the RoI when the camera changes the viewing angle. In order to reduce the manual work for selecting the RoI, many automatic road recognition methods have been proposed [4]. Some methods try to use vehicle motion information to segment the frame into active and inactive traffic regions. They estimate the road region by generating a map for active traffic region based on the trajectories or foreground masks of the moving vehicles. This kind of approaches requires a sufficient number of vehicles to pass along the road. Therefore, the initial time required to gather the required information can vary based on the traffic flow. Some approaches use single images and and try to fit linear or polynomial equations to the straight or curvy road boundaries and lane marks. These methods perform better in case of in-vehicle cameras where the vanishing point is easier to estimate and they are limited to well-structured roads with visible and distinguishable sign-lines which is not always the case. Some methods try to estimate the road boundaries by extracting low-level image features (e.g. color, edge, texture). These methods are usually based on single images and low-level features analysis to classify the pixels or groups of pixels into road and non-road regions. They do not consider the structure or boundaries of the road and only tend to estimate the road area based on the color ([5], [6]), edge ([7], [6], [8]) and texture([9], [10]) of the road surface. Neural networks are also used for road recognition [11], [12]. This kind of supervised learning methods require numerous labeled data for the training process, which is hard to achieve.

In this paper, we propose a statistical modeling method to recognize the road regions automatically in the traffic video analysis. First, we introduce a temporal feature guided statistical modeling method to build the road model. We use the temporal features in videos as a guidance to automatically extract some sample data from the estimated background image. This sample data set mainly contains the features of the road, but also contains some other features. We build a Gaussian mixture model using this data set and further prune the model to get a statistical model for the road. Second, we propose a road recognition method, which can detect the road regions in a video frame. The detected road regions can be used as the RoI for traffic video analysis tasks. In the end, we use some real traffic video sequences from the New Jersey Department of Transportation (NJDOT) to evaluate the performance of our proposed method. The experimental results show that our method is able to detect the road regions accurately and robustly.

978-1-7281-8213-1/20/$31.00 ©2020 IEEE

II. Related Work

Video surveillance cameras are widely deployed in the recent world [13]. How to analyze surveillance videos becomes a very important topic. Traffic surveillance cameras are a kind of most widely seen cameras. Automatic road recognition is an important task in analyzing traffic surveillance videos. Numerous approaches have been taken in order to segment the road region automatically.

The accumulation of the motion trajectories is widely used in road recognition. Stewart et al. [14] distinguish between active and inactive areas of the scene based upon accumulating a map of significant scene changes. Melo et al. [15] model the motion trajectories of tracked vehicles by low-degree polynomials and use a K-means clustering technique on the coefficient space to obtain approximate lane centers. Lee et al. [16] generate a roadway mask image by accumulating moving parts in a difference image between two consecutive input frames and then identify a center line of the roadway to separate two directions of the traffic. These approaches depend on the performance of the foreground detection and the tracking methods which can be affected by the quality of video, changes in illumination, traffic density, etc. Our proposed method utilize the foreground detection result as a guidance to build the statistical model, greatly reduces the inaccuracy caused by the foreground detection method.

Road markings and the lane markings are the most commonly used features for road recognition [17]. Wang et al. [18] introduce an algorithm called CHEVP to initialize a B-spline SNAKE algorithm and use the resulting B-spline curve to represent a curved road. Zhou et al. estimate the lane model parameters (e.g. starting position, orientation, lane width, etc) and generate several lane model candidates and match the best fitted lane model. Aly [19] proposes a real-time algorithm for detecting lane markings in urban streets by taking a top view of the road image, filtering with Gaussian kernels, line detection with Hough transformation, and a new RANSAC spline fitting approach. Kong et al. [20] use the OCR feature to estimate the vanishing point with a clustering method for road recognition. The vanishing point is estimated based on Gabor filters used to compute the dominant texture orientation at each pixel and a new edge detection technique to extract the road boundaries. Son et al. [21] propose a real-time lane detection method to deal with illumination variation to use in lane departure warning system. After using the lane color properties to detect candidates for lane markers a clustering method is applied to find the main lane. Helala et al. [22] segment the road into a number of superpixel regions and use the contours of these regions to generate several edges which are grouped into different clusters and the cluster with the highest confidence score is chosen as the road boundary. However, numerous roads have poor quality that the markings are not clear or missing. It becomes difficult to locate the road when the road quality is poor or the resolution of the video is low.

Most recent studies try to use a combination of low-level and high-level features to deal with the road recognition problems, especially to overcome the effects of illumination changes and strong shadows. Wang et al. [23] introduce a close to realize a road recognition method based on illumination invariant image and quadratic estimation. After extracting an illumination invariant image, a manual triangular road region is used as the color sample to analyze the illumination invariant image and obtain probability maps. The combined probability map is resettled based on histogram analysis, and the road region is estimated for the first time. Then the effective road boundary is extracted after analyzing the gradient image by the estimated road region and the final more accurate road region is obtained. This method follows the assumption in [20] and [24] to use a manual triangular region that is approximated as the initial road estimation model. Tong et al. [25] use simple statistics to propose effective projection angle calculation methods in logarithmic domain to extract the intrinsic images of roads in order to weaken the shadow effect and eliminate the impact of the direction of camera features. This method follows a similar approach based on [26] to use a prior triangle region to sample the color of road region. Li et al. [27] propose a road region extraction algorithm based on vanishing point location. The spatial structure of the road is estimated and uses color and edge features of the intrinsic image extracted based on regression analysis.

Our proposed method uses the result of a foreground detection method as guidance to build a statistical road model, and further applies a discriminant function to segment the road regions in the video. Our proposed pruned mixture model is able to correctly segment the shadowed road regions and poor quality roads.

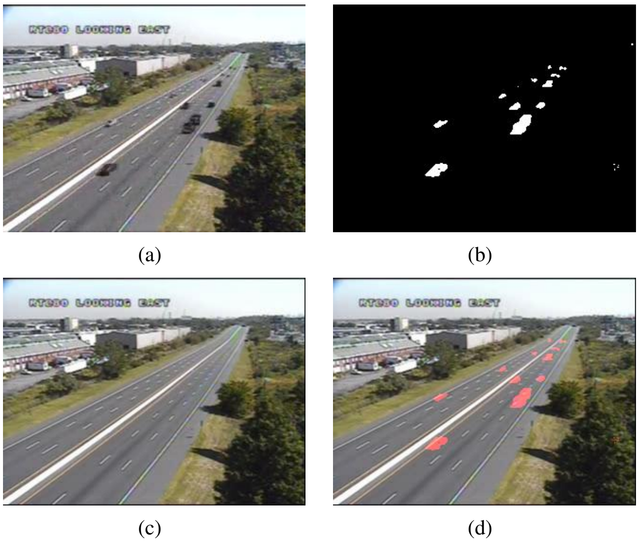
III. A Temporal Feature Guided Statistical Modeling Method for Road Recognition

The RoI selection is a widely used pre-processing technique of many video analysis methods. Manually selecting the RoI is a complex and tiresome task for human beings. Therefore, we propose an statistical modeling method for road recognition, which can detect the region of road automatically without any manual intervention. Our proposed method mainly has two major contributions: (i) The new temporal feature guided statistical modeling method can build the model without any label, which can reduce numerous manual work. (ii) The novel road recognition method can automatically segment the road region as the RoI for traffic video analysis.

A. The New Road Model Estimation Method

When building the statistical model for a class of objects, one common approach is to use some training data that has been labeled as this class to estimate the probability density function. This labeling work may require numerous effort and time. Instead of using manually labeled data, we propose a temporal feature guided model estimation method, which can extract a sample data set from the video based on the temporal features.

In traffic videos, the region we are interested in is the road, which always has moving objects on it. One important information the moving objects can provide is the temporal information. In order to utilize the temporal feature, we apply a foreground object detection method to segment the moving foreground objects and estimate the static background [28], [29]. The foreground detection method is able to detect the areas where have moving objects, and estimate a static background image that does not contain the moving objects.



(a) (b) (c) (d)

Fig. 1: Fig. 1 (a) shows a video frame from an NJDOT traffic video. Fig. 1 (b) displays a binary foreground mask, where white pixels represents the moving foreground objects. Fig. 1 (c) shows the estimated background image. Fig. 1 (d) shows the corresponding regions of the moving foreground objects projected on the background image in red color.

Fig. 1 (a) shows a video frame from an NJDOT traffic video. Fig. 1 (b) displays a binary foreground mask, where white pixels represents the moving foreground objects. Fig. 1 (c) shows the estimated background image. Fig. 1 (d) shows the corresponding regions of the moving foreground objects projected on the background image in red color.

By projecting the the moving foreground mask on the background image, we can get some regions, which contains the temporal features in the orginal video frame. As shown in Fig. 1 (d), the red regions represents the projection of the foreground mask. We can see that most of these areas are the road regions. Therefore, we can extract the feature vectors in these regions from the background image to build the road model.

For each video, we use the first K frames to build the model. Suppose $\mathbf{X} = \{\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}\}$ are the feature vectors we extracted from the K frames. As we only extract features from the regions corresponding to the foreground mask, $\mathbf{X}$ mainly contains the features of the road at different locations and different time. However, some other features are still included in $\mathbf{X}$ due to the noises of the binary foreground mask, or overlapping caused by the viewing angle of the camera. We can use a Gaussian mixture model to estimate the distribution of sample set $\mathbf{X}$ as follows:

$$p\left(\vec{x}|\mathbf{X}\right) = \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\vec{x}; \vec{\mu}_m, \sigma_m^2 I\right) \tag{1}$$

where M is the number of components in the Gaussian mixture model, $\mathcal{N}\left(\vec{x}; \vec{\mu}_1, \sigma_1^2 I\right), \ldots, \mathcal{N}\left(\vec{x}; \vec{\mu}_M, \sigma_M^2 I\right)$ are the Gaussian components. α_m is the weight of the m_{th} Gaussian components, and the summation of $\alpha_1, \ldots, \alpha_M$ is one. The Gaussian components are sorted in descending order according to value of α.

Because the foreground detection result is not 100% accurate, the sample data we used to build the model is noisy. Some non-road feature may also involved in the sample data set. Therefore, we need to prune the Gaussian mixture model in order to get the road model. As we know, the majority of the sample set $\mathbf{X}$ is the feature of the road. The probability of the road features in the sample set is much higher than that of the noises. Therefore, the Gaussians with large weights can be used to describe the road. We select to use the first K Gaussians in $p\left(\vec{x}|\mathbf{X}\right)$ as the road model, which is defined as follows:

$$p\left(\vec{x}|\mathbf{X}, Road\right) = \frac{\sum_{k=1}^{K} \alpha_k \mathcal{N}\left(\vec{x}; \vec{\mu}_k, \sigma_k^2 I\right)}{\sum_{k=1}^{K} \alpha_k} \tag{2}$$

K is defined as:

$$K = \arg\min_{k} \left(\sum_{m=1}^{k} \alpha_m > (1 - \mathcal{T}) \right) \tag{3}$$

where $\mathcal{T}$ is a threshold depends on the portion of the non-road features in the sample set $\mathbf{X}$. For example, if the foreground mask is noisy, we should select a high $\mathcal{T}$ value.

B. The Novel Road Recognition Method

In this section, we introduce a novel road recognition method using the statistical road model. As we know, the feature of the road is relatively simple. Most of the road regions looks similar. The road model is based on Gaussian mixture model, which has peaks at several feature points with highest probability. If the feature vector of a pixel is close to any of these peaks, it would have a higher probability to be a road pixel. Otherwise it is not a road pixel. By using this property of the road model, we proposed a discrimination function to classify the feature vector ($\vec{x}$ of each pixel into a road class and a non-road class. Suppose the road model contains K Gaussian distributions. The discrimination functions are defined as follows:

$$R(\vec{x}) = \begin{cases} Road, & if\ \sum_{m=1}^{K} C(\vec{x})_m > 0 \\ Non-road, & otherwise \end{cases} \tag{4}$$

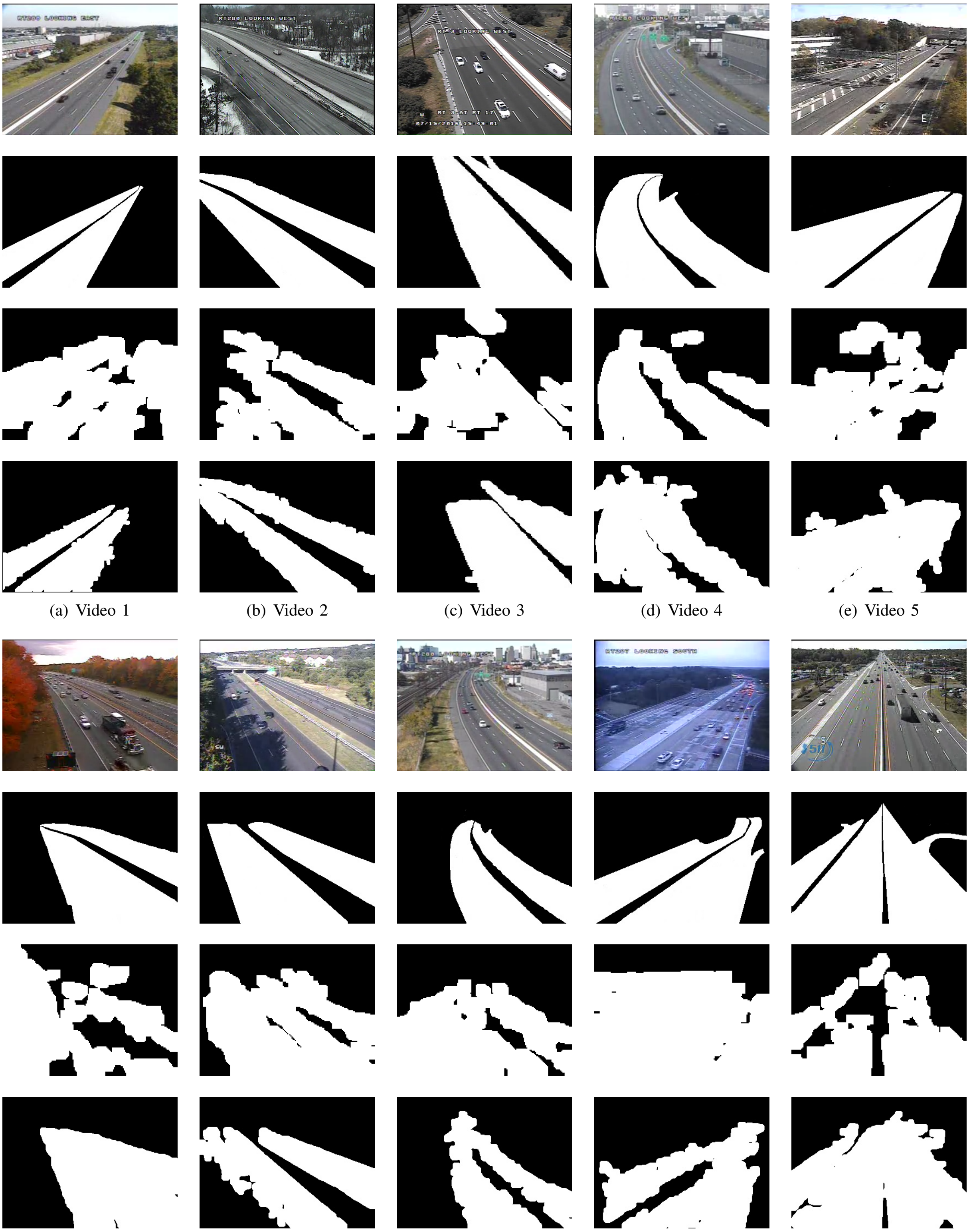
$$C(\vec{x})_m = \begin{cases} 1, & if\ D(\vec{x})_m < 0 \\ 0, & otherwise \end{cases} \tag{5}$$

$$D(\vec{x})_m = (\vec{x} - \vec{\mu}_m)^T (\vec{x} - \vec{\mu}_m) - \sigma_m^2 \tag{6}$$

where $m \in \{1, 2, \ldots, K\}$.

We apply this discrimination function to every pixel in the estimated background image. We can build a road mask and assign every pixel classified as road with value 255, all the other pixels with value 0. Then we can get a binary road mask showing the road pixels.

As we know, the road region is always a large continues region in traffic surveillance videos. However, some pixels on the road may have abnormal features that can not be described by the road model, such as damaged areas, shadows, lane marks, etc. The miss classification of these pixels may cause the road mask has some holes in the road region. In addition, some non-road pixels may be detected as road because they



(a) Video 1 (b) Video 2 (c) Video 3 (d) Video 4 (e) Video 5

(f) Video 6 (g) Video 7 (h) Video 8 (i) Video 9 (j) Video 10

Fig. 2: The road recognition results. The first and the fifth rows display one video frame from an NJDOT traffic video. The second and the sixth rows show the ground truth road regions. The third and the seventh rows present the road recognition result of UFL-HS method. The fourth and the eighth rows show the road region detected by our proposed method.

have similar features as the road. This will cause some noises

outside the road region. In order to solve these issues, we apply

TABLE I: The quantitative performance of the proposed method

Method	Video #	1	2	3	4	5	6	7	8	9	10	Average
UFL-HS Method	Precision	0.41	0.73	0.62	0.82	0.68	0.43	0.71	0.53	0.57	0.91	0.64
	Recall	0.89	0.79	0.68	0.86	0.81	0.62	0.93	0.91	0.95	0.78	0.82
	F-Score	0.56	0.76	0.65	0.84	0.74	0.50	0.80	0.67	0.71	0.84	0.72
Proposed Method	Precision	0.94	0.96	0.93	0.85	0.88	0.90	0.95	0.91	0.90	0.89	0.91
	Recall	0.89	0.86	0.78	0.91	0.94	0.98	0.85	0.92	0.88	0.97	0.90
	F-Score	0.91	0.91	0.85	0.88	0.91	0.94	0.90	0.91	0.89	0.92	**0.90**

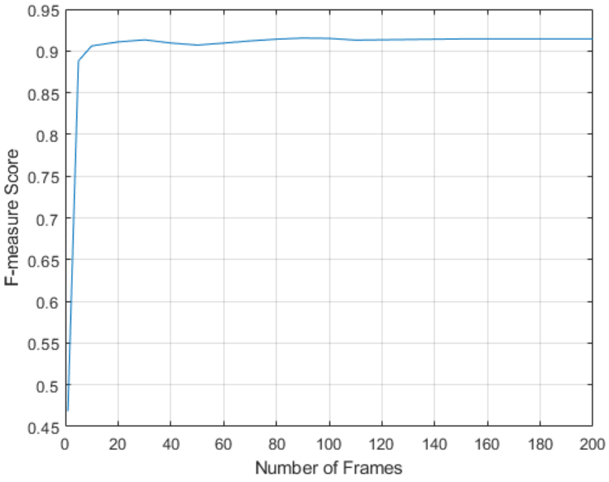


Fig. 3: The F-measure score of our proposed road recognition method using different number of training frames.

a morphological operator on the road mask to further enhance the road recognition result. The morphological operation is defined as follows:

$$R'_{mask} = (R_{mask} \ominus E) \oplus D \qquad (7)$$

where R_{mask} is the road mask, $\ominus$ is the erosion operator, E is the erosion template, $\oplus$ is the dilation operator, and D is the dilation template.

IV. Experiments

In this section, we show some experimental results to evaluate our statistical modeling method for the road recognition. The data set we use contains the real traffic surveillance videos from the New Jersey Department of Transportation (NJDOT). To ensure the diversity of the videos, this data set includes ten video sequences with several kinds of resolutions and frame rates, various weather conditions, and different illumination conditions. One frame of each video is displayed in the first and the fifth rows in Fig. 2. The second and the sixth rows in Fig. 2 show the ground truth road region masks. The third and the seventh rows in Fig. 2 present the road recognition result of UFL-HS method [30]. The fourth and the eighth rows in Fig. 2 show the road region detected by our proposed method. We can see from Fig. 2, our statistical modeling method can detect the road regions in all the videos accurately.

The feature vector $\vec{x}$ we used in the experiment is the histofram of oridented gradients (HOG) feature [31] calculated from a 4×4 cell surrounding each pixel. The number of component in the Gaussian mixture model M is 3. The number of frames used for building the model is 50.

We further compare our method with the UFL-HS method [30] quantitatively. The precision, recall and the F-measure score are popular metrics used to evaluate detection performance, which are defined as follows:

$$Precision = \frac{TP}{TP + FP} \qquad (8)$$

$$Recall = \frac{TP}{TP + FN} \qquad (9)$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (10)$$

where TP, FP and FN represent the number of true positive, false positive, and false negative detections of the road pixel. In Table. I, we show the quantitative results of our proposed method. We can see the average accuracy of the road region detected by our method is our 90%, which is good enough for it to be used as the RoI.

We further investigate the influence of the number of frames used for the model estimation. We change the number of frames that used for building the statistical model and calculate the road recognition accuracy. As shown in Fig. 3, the F-measure score is stable around 0.9 when the number of frames is over 20. Hence our proposed method does not need to use a large number of frames to build the model, the model estimation process can be fast enough to perform as a pre-processing step of video analysis.

V. Conclusions

In this paper, we have proposed a statistical modeling road recognition method, which switches the info-communications between the ROI selection and video analysis systems from an inter-cognitive communication mode to an intra-cognitive communication mode. The novel road recognition method uses the temporal features in videos instead of manual labeling to generate sample data for statistical modeling and recognizes the road regions which can be used as ROIs in video analysis systems. Our proposed method on one hand improves the road detection accuracy compared to the state-of-the-art method, on the other hand, provides another option for the cognitive info-communications between the ROI selection and video analysis systems. The experimental results using the real traffic video sequences from NJDOT verify the robustness and accuracy of our proposed method.

VI. Acknowledgments

This paper is partially supported by the NSF grant 1647170.

REFERENCES

[1] R. Brinkmann, *The art and science of digital compositing: Techniques for visual effects, animation and motion graphics*. Morgan Kaufmann, 2008.

[2] P. Baranyi and Á. Csapó, “Definition and synergies of cognitive infocommunications,” *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

[3] P. Baranyi, A. Csapo, and G. Sallai, *Cognitive Infocommunications (CogInfoCom)*. Springer, 2015.

[4] V. H. Mistry and R. Makwana, “Survey: Vision based road detection techniques,” *Int. J. Comput. Sci. Inf. Technol*, vol. 5, pp. 4741–4747, 2014.

[5] C. Tan, T. Hong, T. Chang, and M. Shneier, “Color model-based real-time learning for road following,” in *2006 IEEE Intelligent Transportation Systems Conference*. IEEE, 2006, pp. 939–944.

[6] J. M. Á. Alvarez and A. M. Lopez, “Road detection based on illuminant invariance,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2010.

[7] Y. Zhou, R. Xu, X. Hu, and Q. Ye, “A robust lane detection and tracking method based on computer vision,” *Measurement science and technology*, vol. 17, no. 4, p. 736, 2006.

[8] M. A. Sotelo, F. J. 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.

[9] Y.-W. Seo and R. R. Rajkumar, “Detection and tracking of boundary of unmarked roads,” in *17th International Conference on Information Fusion (FUSION)*. IEEE, 2014, pp. 1–6.

[10] S. Zhou, J. Gong, G. Xiong, H. Chen, and K. Iagnemma, “Road detection using support vector machine based on online learning and evaluation,” in *2010 IEEE Intelligent Vehicles Symposium*. IEEE, 2010, pp. 256–261.

[11] A. Narayan, E. Tuci, F. Labrosse, and M. H. M. Alkilabi, “Road detection using convolutional neural networks,” in *Artificial Life Conference Proceedings 14*. MIT Press, 2017, pp. 314–321.

[12] Y. Lyu, L. Bai, and X. Huang, “Road segmentation using cnn and distributed lstm,” in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2019, pp. 1–5.

[13] D. Palivcova, M. Macik, and Z. Mikovec, “Susy: Surveillance system for hospitals,” in *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2017, pp. 000 131–000 136.

[14] B. Stewart, I. Reading, M. Thomson, T. Binnie, K. Dickinson, and C. Wan, “Adaptive lane finding in road traffic image analysis,” 1994.

[15] J. Melo, A. Naftel, A. Bernardino, and J. Santos-Victor, “Detection and classification of highway lanes using vehicle motion trajectories,” *IEEE Transactions on intelligent transportation systems*, vol. 7, no. 2, pp. 188–200, 2006.

[16] W. Lee and B. Ran, “Bidirectional roadway detection for traffic surveillance using online cctv videos,” in *2006 IEEE Intelligent Transportation Systems Conference*. IEEE, 2006, pp. 1556–1561.

[17] A. B. Hillel, R. Lerner, D. Levi, and G. Raz, “Recent progress in road and lane detection: a survey,” *Machine vision and applications*, vol. 25, no. 3, pp. 727–745, 2014.

[18] Y. Wang, E. K. Teoh, and D. Shen, “Lane detection and tracking using b-snake,” *Image and Vision computing*, vol. 22, no. 4, pp. 269–280, 2004.

[19] M. Aly, “Real time detection of lane markers in urban streets,” in *2008 IEEE Intelligent Vehicles Symposium*. IEEE, 2008, pp. 7–12.

[20] H. Kong, J.-Y. Audibert, and J. Ponce, “Vanishing point detection for road detection,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 96–103.

[21] J. Son, H. Yoo, S. Kim, and K. Sohn, “Real-time illumination invariant lane detection for lane departure warning system,” *Expert Systems with Applications*, vol. 42, no. 4, pp. 1816–1824, 2015.

[22] M. A. Helala, K. Q. Pu, and F. Z. Qureshi, “Road boundary detection in challenging scenarios,” in *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*. IEEE, 2012, pp. 428–433.

[23] E. Wang, Y. Li, A. Sun, H. Gao, J. Yang, and Z. Fang, “Road detection based on illuminant invariance and quadratic estimation,” *Optik*, vol. 185, pp. 672–684, 2019.

[24] Y. Li, W. Ding, X. Zhang, and Z. Ju, “Road detection algorithm for autonomous navigation systems based on dark channel prior and vanishing point in complex road scenes,” *Robotics and Autonomous Systems*, vol. 85, pp. 1–11, 2016.

[25] G. Tong, Y. Li, A. Sun, and Y. Wang, “Shadow effect weakening based on intrinsic image extraction with effective projection of logarithmic domain for road scene,” *Signal, Image and Video Processing*, pp. 1–9, 2019.

[26] E. Wang, A. Sun, Y. Li, X. Hou, and Y. Zhu, “Fast vanishing point detection method based on road border region estimation,” *IET Image Processing*, vol. 12, no. 3, pp. 361–373, 2017.

[27] Y. Li, G. Tong, A. Sun, and W. Ding, “Road extraction algorithm based on intrinsic image and vanishing point for unstructured road image,” *Robotics and Autonomous Systems*, vol. 109, pp. 86–96, 2018.

[28] H. Shi and C. Liu, “A new global foreground modeling and local background modeling method for video analysis,” in *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer, 2018, pp. 49–63.

[29] ——, “A new foreground segmentation method for video analysis in different color spaces,” in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 2899–2904.

[30] J. Zhang, S. Xia, K. Lu, H. Pan, and A. K. Qin, “Robust road detection from a single image,” in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 859–864.

[31] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” 2005.