2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)

Automatic Road Detection in Traffic Videos

Hadi Ghahremannezhad
Computer Science Department
New Jersey Institute of Technology
Newark, New Jersey 07102
Email: hg255@njit.edu

Hang Shi
Computer Science Department
New Jersey Institute of Technology
Newark, New Jersey 07102
Email: hs328@njit.edu

Chengjun Liu
Computer Science Department
New Jersey Institute of Technology
Newark, New Jersey 07102
Email: cliu@njit.edu

***Abstract*—Automatic road detection is a challenging and representative computer vision problem due to a wide range of illumination variations and weather conditions in real traffic. This paper presents a novel real-time road detection method that is able to accurately and robustly extract the road region in real traffic videos under adverse illumination and weather conditions. Specifically, the innovative global foreground modeling (GFM) method is first applied to robustly model the ever-changing background in the traffic as well as to accurately detect the regions of the moving objects, namely the vehicles on the road. Note that the regions of the moving vehicles are reasonably assumed to be the road regions, which are then utilized to generate in total seven probability maps. In particular, four of these maps are derived using the color values in the RGB and HSV color spaces. Two additional probability maps are calculated from the two normalized histograms corresponding to the road and the non-road pixels in the RGB and grayscale color spaces, respectively. The last probability map is computed from the edges detected by the Canny edge detector and the regions located by the flood-fill algorithm. Finally, a novel automatic road detection method, which integrates these seven masks based on their probability values, defines a final probability mask for accurate and robust road detection in video.**

I. Introduction

Region of interest (RoI) determination is one of the most important and fundamental pre-processing steps in many image and video processing applications. A well-defined region of interest contains all the important location of the image while excluding the unnecessary regions from the processed data in the tasks of image and video analysis. Lowering the amount of required computational resources, increasing the processing speed, and reducing faulty results are some of the benefits of determining an RoI. The focus of this study is automatic RoI determination in traffic videos which is mostly associated with the road region where the vehicles are located. Road region recognition is a crucial step in many computer vision applications such as self-driving vehicles, intelligent driver assistant systems, traffic surveillance, and navigation systems.

There have been many studies addressing the issue of vision-based road recognition in recent years in applications regarding in-vehicle cameras [1]–[3] and traffic video surveillance [4]–[7]. Most recently, convolutional neural networks have attracted a lot of attention in computer vision applications including road segmentation [8]–[13]. However, the need for large amount of training data and computational resources along with lack of sufficient generalization ability, makes it difficult to apply these methods in real-world applications.

In this paper, we propose an adaptive road recognition method that extracts the road location from single frames in a traffic video sequence and further updates and refines the estimated road region as more video frames are processed. No assumption about the structure of the road is made and therefore, this method can be used for structured and unstructured road scenarios. The locations of moving vehicles are appropriately assumed to associate with the roadway region and they are utilized as color samples to estimate the location of road pixels. A novel foreground segmentation technique [14] based on Gaussian mixture models is applied in order to detect the moving vehicles and subtract the stable background. The pixel values of the background image at the corresponding location of the vehicles are utilized as initial road samples and several road probability maps are generated. The extracted probability values are then combined in order to estimate a more accurate road region map which is further refined by using the aggregated foreground mask.

The remainder of this paper is organized as follows. Section II describes the proposed road detection method in details. In section II-A an approach is discussed to define the initial road samples based on the location of moving vehicles. Section II-B contains details on generating several road probability maps which are further combined and refined by the approach narrated in section II-C. The performance of the proposed method is evaluated by using real traffic videos in section III, and we conclude the paper in section IV.

II. A new automatic method for road region extraction

Manually determining the region of interest, which in traffic video analysis refers to the road region, is an exhaustive and time-consuming task for human agents. Here, a fully automatic approach for road segmentation and RoI determination is proposed to reduce the manual effort. The proposed method can be performed in real-time and is adaptive to camera view changes and various illumination scenarios. The only assumption made is about the location of the vehicles which are assumed to move mostly along the road region. Our proposed method has mainly two contributions: (i) The new road probability estimation method can generate a reliable road map from the initial frames of the video without the need to

978-1-6654-1485-2/20/$31.00 ©2020 IEEE
DOI 10.1109/ISPA-BDCloud-SocialCom-SustainCom51426.2020.00123

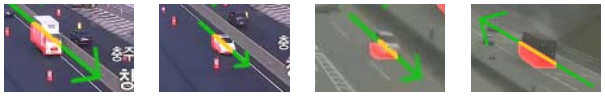


Fig. 1: Sampling the road pixels from the background image based on the direction of moving vehicles in order to avoid sampling non-road pixels. The red color indicated the location of the sampled road pixels.

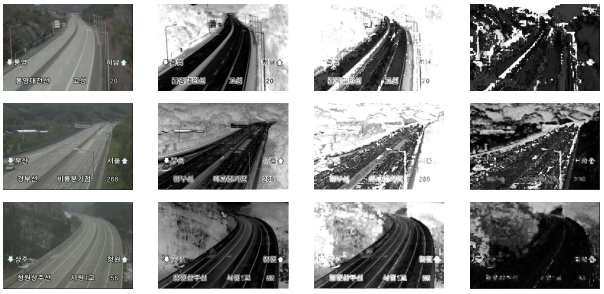
wait for many vehicles to pass along the road region. (ii) The novel road segmentation method can automatically refine the initial road map and find the region of interest to use in traffic surveillance video analysis tasks.

A. Selection of the initial road samples

In case of applications with an on-board camera system, initial road samples are usually taken from a triangular area in front of the vehicle. In contrast, in applications with a stationary camera overlooking the roadway, the initial road samples can be extracted based on the location of moving vehicles. The further steps for road segmentation based on the initial samples can be commonly used among applications of traffic surveillance and self-driving vehicles. The focus of this study is on automatic RoI determination in traffic surveillance videos. However, our proposed feature extraction and classification approach can work for road segmentation in self-driving vehicles as well.

In order to obtain an estimate of the road region during the initial frames of the video, first we attempt to detect the vehicles and segment them from the still background. The global foreground modeling (GFM) method introduced in [14], [15] is utilized to detect the location of the moving vehicles and to subtract the stationary background image from the video frames. The GFM foreground segmentation approach is chosen due to its ability to quickly subtract the background in a video captured by a stationary camera. Also, the GFM method is robust in dealing with stopped vehicles which are continuously detected as foreground and therefore separated from the background image. The road estimation method is applied on the subtracted background with the assumption that most vehicles pass along the roadway. The corresponding location of the moving and stopped vehicles in the background image is considered to be samples of the road region which are in turn utilized to estimate the probability of all background pixels. The generated probability maps are further used to classify the pixels into road and non-road in order to segment the road region from other areas and determine the RoI based on the extracted road map.

The selected pixels for road samples should be exclusively from the road region in order to obtain a good estimation of road pixel-values. In many intelligent vehicle systems such as automatic driving, and advanced driver assistance systems where the field of view is similar to that of the driver's, the road region priori is approximated as a triangular region at the mid-bottom of the frame [16]–[18]. In case of



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

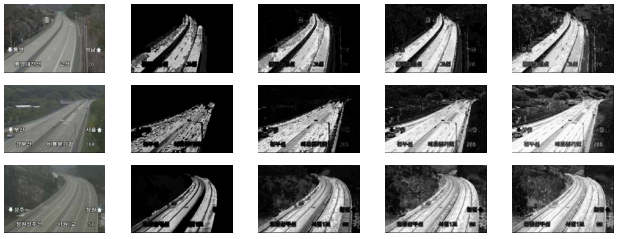
Fig. 2: Extracting the auxiliary road region probability maps using difference images. (a) The background image. (b), (c), (c) are the gray-scale, color, and hue difference images, respectively.

traffic surveillance videos where the cameras are overlooking the road, there can be no initial assumption of the road's location without any observation of the images. Here, a valid assumption is made that most of the pixels in the background image with locations corresponding to those of the vehicles in the foreground mask belong to the roadway region. However, due to the variety of camera view angles, different sizes of vehicles, and occasional movements in the non-road regions, some of the pixels of the foreground mask can belong to the areas outside of the road. Therefore, the vehicles are tracked and after a few number of initial frames, the foreground mask is filtered by taking into account the moving direction, track lifetime, and the displacement vector size of each vehicle. So that it is possible to comply with real-time constraints of the traffic management systems, a fast multi-object tracking method [19] is applied.

In order to obtain a mask containing pixels that represent road samples Ω_{rsm}, only the foreground mask of vehicles with sizable movement and long enough tracking life-time are considered. The moving direction of each vehicle is estimated and updated as follows in each sequence of f frames:

$$\begin{aligned} v_x &= x_{m_2} - x_{m_1} \\ v_y &= y_{m_2} - y_{m_1} \\ d_i &= arctan(v_y, v_x) \\ m_{v_i} &= \sqrt{v_x^2 + v_y^2} \end{aligned} \tag{1}$$

where v_x and v_y are the components of the velocity vector, x_{m_2} and y_{m_2} are the average x and y values of the blob centroid in the most recent $f/2$ frames, x_{m_1} and y_{m_1} are the average x and y values of the blob centroid in the remaining $f/2$ frames, d_i is the estimated direction of the vehicle i, and m_{v_i} is the estimated magnitude of the vehicle i, respectively. The filtered foreground mask of each vehicle is then cropped with regards to its moving direction so that only the part that corresponds to the road region is added to the Ω_{rsm} mask. Figure 1 illustrates some examples of the road sampling strategy which helps avoid including non-road regions in the



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

Fig. 3: Extracting the auxiliary road region probability maps using difference images. (a) The background image. (b), (c), (d), (e) are the extracted probability maps P_G, P_C, P_H, and P_S, respectively.

Ω_{rsm} at the boundaries of the roadway. The road samples are accumulated throughout the video and the Ω_{rsm} mask will cover more parts of the road when more vehicles pass along the roadway.

B. Road region probability map extraction

Generating a single probability map that represents the roadway region in all cases is rather difficult due to the variety of illumination, texture, color and other visual conditions. Therefore, generating multiple probability maps and merging them helps obtaining a more reliable probability distribution for classifying the pixels into road and non-road regions. In this section, multiple approaches are taken in order to generate a number of probability maps using low-level features, e.g., color, edge, and temporal features. The generated probability maps are further combined together to obtain a binary classification mask which is in turn refined by the accumulative foreground mask as the number of passing vehicles increases.

1) Extraction of probability maps based on difference images: One approach to estimate the road probability of the pixels is to compare the pixel's value to the average value of the initially selected road samples in Ω_{rsm}. Similar to the approach used in [20], [21] the gray-scale image G^* of background is first smoothed by applying a Gaussian convolution kernel of size 3×3 to reduce the noise effect. Then the absolute difference between the mean value $\bar{G^*_{rsm}}$ of the grayscale image in the location of Ω_{rsm} and each pixel in the smoothed grayscale image is utilized to obtain a gray-scale difference image G. A similar process is carried out on the three channels of the smoothed background image and the three outputs are added together to obtain another difference image C based on the color input. In traffic scenes where the roadway is considerably different in color from the surrounding area the hue channel of HSV color space can be a distinguishable factor in segmenting the road pixels from the image, especially at the boundaries of the road. The background image is also converted to HSV color space and the hue channel is utilized to acquire a difference image H through a similar process. Figure 2 illustrates sample difference images obtained from real traffic video data.

Lower values in the difference images correspond to the parts of the image that are closer to the average value of the road pixels in Ω_{rsm} and have a higher probability of belonging to the road region. Therefore, the probability value of each pixel should be inversely proportional to the corresponding pixel in the difference image. Probability maps can be estimated accordingly based on the difference images obtained so far in which the probability of each pixel is calculated as follows:

$$P'_K(p_i) = \frac{1 - K(p_i)}{max(K(p_i)|p_i \in K)} \tag{2}$$

where $i = 1...N$ is the pixel index, $K \in \{G, C, H\}$ refers to each difference image, and $P'_K(p_i)$ is the probability of the pixel p_i belonging to the road region in the difference image K. In order to normalize the brightness and increase the probability contrast of the probability maps, their histograms are normalized to obtain an approximation of the probability density function, and the normalized histograms are equalized as follows:

$$\begin{aligned} H'_{n,P'_K} &= \sum_{0 \le m < n} H_{P'_K}(m) \\ P_K(p_i) &= H'_{P'_K}(P'_K(p_i)) \end{aligned} \tag{3}$$

where $i = 1...N$ is the pixel index, $K \in \{G, C, H\}$ represents each difference image, $H_{P'_K}$ and $H'_{P'_K}$ are the normalized histogram and the integral histogram of probability map P'_K respectively, and P_K refers to the equalized histogram of each probability map.

The pixels representing the road region in traffic videos usually have a close value in most parts of the roadway contained in the frame and the road samples represent a high percentage of the road pixels. Therefore, the standard deviation is usually assumed to have a relatively small interval with a high level of confidence. The further the pixel values in G are from the standard deviation of the pixels in the road sample mask Ω_{rsm}, the probability of belonging to the road region should drop. Considering the standard deviation of the road samples, another probability map can be obtained as follows that specifically favors the pixels that are close to the road samples:

$$\begin{aligned} \alpha(p_i) &= max(0, sgn(G(p_i) - \sigma_{rsm})) \\ P_S(p_i) &= 1 - \alpha(p_i)[\frac{G(p_i)}{k\sigma_{rsm}} + \frac{1}{k^2}], k - 1 \le \frac{G(p_i)}{\sigma_{rsm}} < k \end{aligned} \tag{4}$$

where $p_i \in G$, $i = 1...N$, σ_{rsm} is the standard deviation of the pixel values in Ω_{rsm} mask of G, k is a natural number in $\{k \in \mathbb{N} | 1 < k \le max(G(p_i) - \sigma_{rsm})\}$, and $P_S(p_i)$ is the resulting probability map. Figure 3 represents the extracted probability maps from the difference images.

2) Extraction of probability maps based on histogram models: Another approach of estimating the road region probability of each frame is to utilize histogram models extracted from the road and non-road samples. Similar to the approaches used in [16], [22], a similarity measure is used in order to generate probability maps that help classify the road and non-road pixels. The non-road samples are taken from the regions

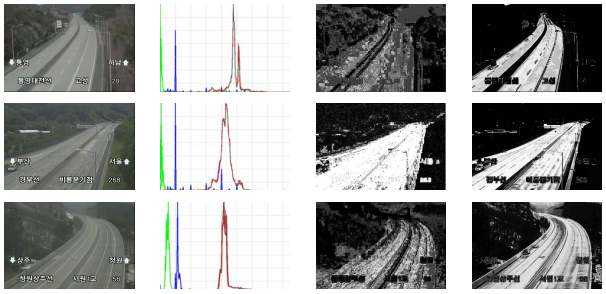
outside of the final estimated road region in the previous frame. The normalized histograms of the blue and green channels of the background image and the gray-scale image G^* are used to estimate probabilities as follows:

$$P_K(p_i) = \frac{N_K^r(K(p_i))}{N_K^r(K(p_i)) + N_K^{nr}(K(p_i))} \tag{5}$$

where $i = 1...N$ is the pixel index, $K \in \{Blue, Green, Gray\}$ refers to the blue and green channels of the background image and the gray-scale image G^*, $N_K^r(K(p_i))$ and $N_K^{nr}(K(p_i))$ are the values of the $K(p_i)th$ bin in the histogram models obtained from the road samples in Ω_{rsm} and non-road samples of the previous frame respectively, and $P_K(p_i)$ is the probability of the pixel p_i belonging to the road region in the image K. Since the histogram models of the red channel and gray-scale of background image are close (as seen in Figure 4(b)), the red-channel histogram is not considered and two probability maps P_{Ghist} and $P_{GBhist} = P_{Green} + P_{Blue}$ are obtained from the gray-scale image G^* and a combination of green and blue channels of the background image, respectively.

3) Extraction of probability maps based on edge information: In many road detection methods [20], [23]–[26] gradient filters are applied in order to differentiate between the road and non-road regions based on the presumed fact that the road region contains considerably less amount of gradient information compared to the surrounding areas. This is usually not the case in traffic surveillance videos where the objects are further from the camera and the edge density is not much higher in the non-road regions. However, the dominant road boundaries create strong edges which can be used along with the location of the vehicles to separate the road region from the surroundings. The Canny edge detection method is applied on the gray-scale difference image G with lower and upper thresholds set to $\tau_l = 0.66 \times M$ and $\tau_h = 1.33 \times M$ respectively, where M is the median luminance of G. Since the geometric distortion caused by the perspective view of the camera lens results in the losing valuable edge information in the regions that are further from the camera. Therefore, the horizontal line can be estimated and considered as a secondary boundary in addition to the background edges in order to avoid including the areas like sky above the vanishing point inside the road region.

In order to avoid the inclusion of non-road pixels as seed points for flood-fill operation, a single block from the colored difference image C located at one of the corner points of each vehicle's surrounding bounding box is chosen as road samples. The selected corner is picked according to the moving direction of each vehicle in order to make sure the sample block is certain to belong entirely to the road region. The pixels in the chosen blocks form a flood seed mask Ω_{fsm} which contains the starting nodes for the procedure of flood-fill algorithm. The extracted edges from the gray-scale difference image G along with the horizontal line are used as boundaries for the flood-fill algorithm with a connectivity value of 4, in order to fill the connected components with a constant value



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

Fig. 4: Extracting the road region probability maps using histogram models. (a) The background image. (b) The histogram plot representing the RGB channels and gray-scale image of the background image. 4(c), (d) are the extracted probability mapa P_{GBhist} and P_{Ghist}, respectively.

in a flood-fill mask image M_F. The maximal lower and upper intensity difference between the currently observed pixel and one of its four nearest neighbors of the same component, or a new seed pixel being added to the component is calculated based on the standard deviation of the colored difference image C as follows:

$$\begin{aligned} m &= \frac{1}{N}\sum_{i=1}^{N} C(p_i) \\ s &= \sqrt{\frac{\sum_{i=1}^{N}(C(p_i) - m)^2}{N}} \\ thr &= max(1, \frac{s}{k}) \end{aligned} \tag{6}$$

where m is the mean value of the colored difference image C, N is the total number of pixels in the background image, p_i is the intensity value of the i-th pixel, k is a pre-defined constant, and thr is the maximal lower or upper intensity difference.

C. Updating and merging the extracted probability maps

The extracted probability maps are updated in order to bring into account the gathered information from all observed frames. As more vehicles pass along different locations of the roadway, the number of pixels in the Ω grows which makes the probability maps of the latest frames more reliable than the initial values. Also, when a pixel repeatedly appears in the foreground mask of the moving vehicles, it is more likely to belong to the road region. Therefore, all probability maps are updated by applying the temporal fusing algorithm at each frame as follows:

$$\begin{aligned} P_K^t(p_i) &= \frac{\sum_{f=1}^{t} w_i^f \times P_K^f(p_i)}{1 + \sum_{f=1}^{t} w_i^f} \\ w_i^f &= \sum_{j=1}^{N} \Omega_M^f(p_j) \end{aligned} \tag{7}$$

where $i = 1...N$ is the pixel index, w_i^f is the weight associated with pixel p_i at frame f, $K \in$

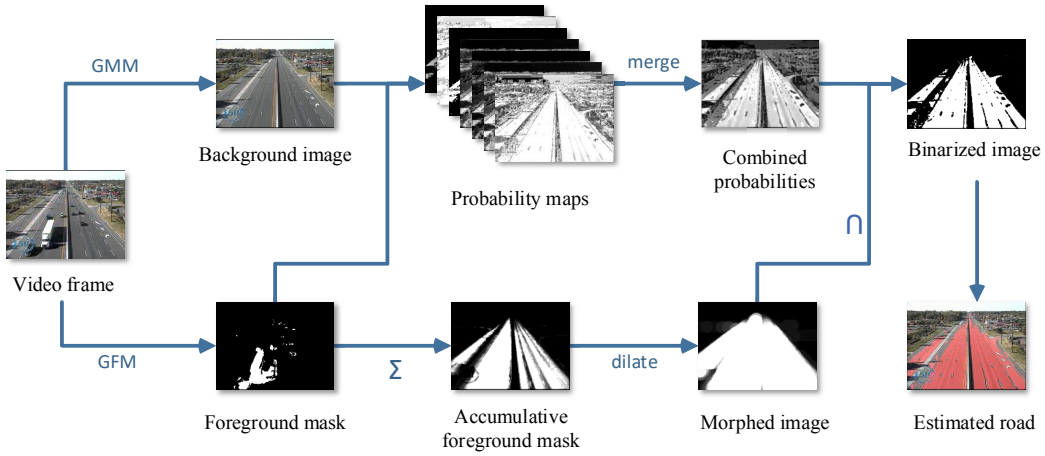


Fig. 5: The process of merging and refining the probability maps. The extracted probability maps are combined and the Otsu's threshold is applied on the result. The non-road pixels that are misclassified as a part of the road region due to similar color values are later filtered out by intersecting the binary image with the accumulative foreground mask.

$\{G, C, H, S, Ghist, GBhist, F\}$ refers to the source of each probability map, $P_K^f(p_i)$ is the probability value of pixel p_i at frame f, $M \in \{rsm, fsm\}$ is the source of the sample mask containing the initial seed points, $\Omega_M^f(p_i) \in \{0, 1\}$ is the value of p_i in the accumulative road sample mask of frame f, N is the total number of pixels in each frame, and $P_K^t(p_i)$ is the updated probability value of pixel p_i.

The updated probability values for each pixel extracted from different sources should be combined with each other, in order to obtain a consensus estimation. If we denote the set of all pixels with $\mathcal{N}$ and the set of extracted probability maps with $\mathcal{K}$, the event R_i specifying whether a pixel $i \in \mathcal{N}$ belongs to the road region or not, can be considered as a Bernoulli random variable $Ber(q_i)$ where $q_i \in [0, 1]$. $R_i = 1$ means i belongs to the road region and $R_i = 0$ means i is a non-road pixel. The set of generated probability maps $\mathcal{K}$, contains several estimations, each of which is drawn from a different source of information. We denote the probability prediction of source j made on pixel i with $p_{i,j} \in [0, 1]$. To solve a probability aggregation problem, we need to design a function $F : ([0, 1])^{|\mathcal{N}|\times|\mathcal{K}|} \rightarrow [0, 1]^{|\mathcal{N}|}$ that takes the predicted probabilities $\{p_{i,j}\}_{i\in\mathcal{N},j\in\mathcal{K}}$ as input and generates an aggregated probability estimation $\hat{q}_i \in [0, 1]$ for each pixel i.

Some simple approaches to aggregate probability predictions are arithmetic mean of the probabilities, median of the probabilities, majority voting, logarithmic opinion pool, and Beta-transformed linear opinion pool. Here, we use weighted mean and median in order to solve the aggregation problem by considering the different degrees of reliability among the generated probability maps and also, taking into account that the aggregated estimation should tend towards the majority opinion in extreme cases of probability predictions. The values of each pixel i in the set $\mathcal{K}$ is sorted and the resulting ordered list $\mathcal{K}' = \{P'_1, ..., P'_m\}$ is utilized to define the weighted median $p'_{i,k}$ such that:

$$\sum_{j=1}^{k-1} w_j \leq 1/2 \quad and \quad \sum_{j=k}^{|\mathcal{K}'|} w_j \leq 1/2 \tag{8}$$

where $j = 1...\mathcal{K}$ is the index of the probability maps and w_j is the weight for each map representing its reliability. Experimental results have shown higher stability of the P_F and P_S probability maps and higher weights are assigned to these source in the aggregation process.

If the values of a pixel in the set of extracted probability maps $\mathcal{K} = \{P_1, ..., P_m\}$ have a large median, it means that the pixel has a high value in most probability maps and therefore, is most likely inside the road region. On the other hand, low median means most predictions contain a low value for a pixel and it most likely belongs to the non-road area. The aggregated probability values are calculated as follows:

$$\hat{q}_i = \begin{cases} \frac{1}{(m-k+1)} \sum_{j=k}^{m} p'_{i,j} & , \text{if } p'_{i,k} > \theta_1 \\ \frac{1}{k} \sum_{j=1}^{k} p'_{i,j} & , \text{if } p'_{i,k} < (1-\theta_1) \\ \frac{1}{2}(p'_{i,k} + \frac{\sum_{j\in\mathcal{K}} w_j p_{i,j}}{\sum_{j\in\mathcal{K}} w_j}) & , \text{otherwise} \end{cases} \tag{9}$$

where $i \in \mathcal{N}$ is a pixel, $p'_{i,j}$ is the probability value of pixel i in the sorted probability set $\mathcal{K}' = \{p'_{i,j}\}_{i\in\mathcal{N},j\in\mathcal{K}'}$, k is the index of the weighted median value $p'_{i,k}$, θ is a pre-defined threshold close to 1, and $\hat{q}_i$ is the aggregated probability value for pixel p_i. The Otsu's threshold [27] is applied on the resulting map in order to filter out the regions with low probability value.

When the intersection between the binary probability mask and the aggregated foreground mask surpasses a threshold, the accumulative foreground mask has covered most of the road pixels after morphological dilation with a size close to the average size of vehicles. As illustrated in fig. 6, the intersection between the accumulative foreground mask and the binary fused probability mask is utilized as the final

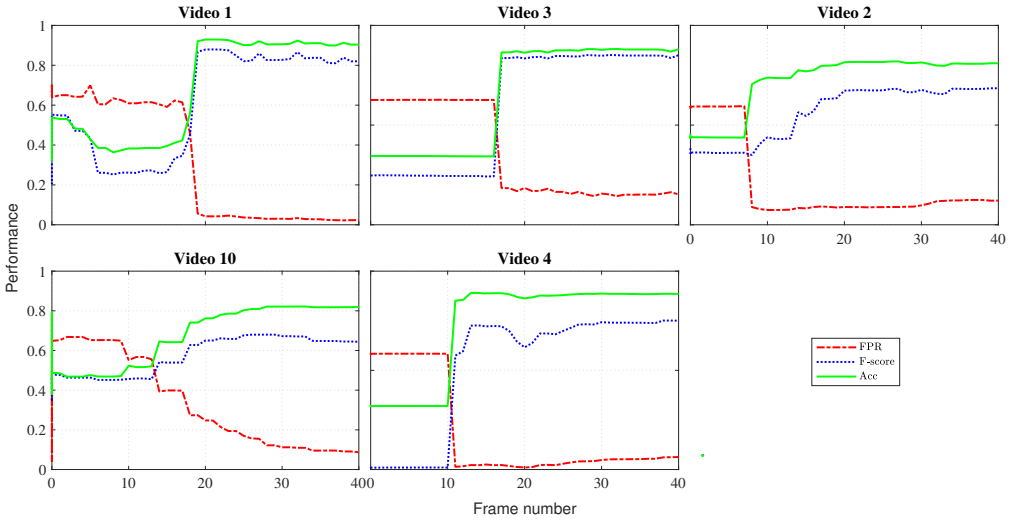


Fig. 6: The F-measure score, accuracy, and false-positive rate of the proposed road extraction method at different frames, tested on some of the sample traffic videos. The sudden improvement in the performance measures happens when the first vehicle is observed in the video sequence and the initial road samples are obtained based on its location.

estimated road region. This way, the possible misclassified non-road regions are removed and the final road map is refined by the exclusion of the over segmentation and leak segmentation errors. In the rare cases of misclassifying the non-road regions as road pixels, the accumulative foreground mask can be applied to remove the misclassified regions from the final road mask while keeping the information about road boundaries. The foreground mask at each frame is obtained by applying the GFM method along with the blob-tracking approach in order to only consider the location of moving vehicles. The foreground blobs are selected by considering their size and moving distance in order to filter out the blobs corresponding to the non-vehicle objects at the non-road regions and slow-moving vehicles. The entire foreground mask of a vehicle is added to the accumulative mask with a weight equal to its moving distance only when the track associated to that vehicle has been inactive for a certain period of time. This way, the vehicles with larger moving distances contribute more to the accumulative foreground mask. The final mask is normalized at each frame as it is divided by the maximum value.

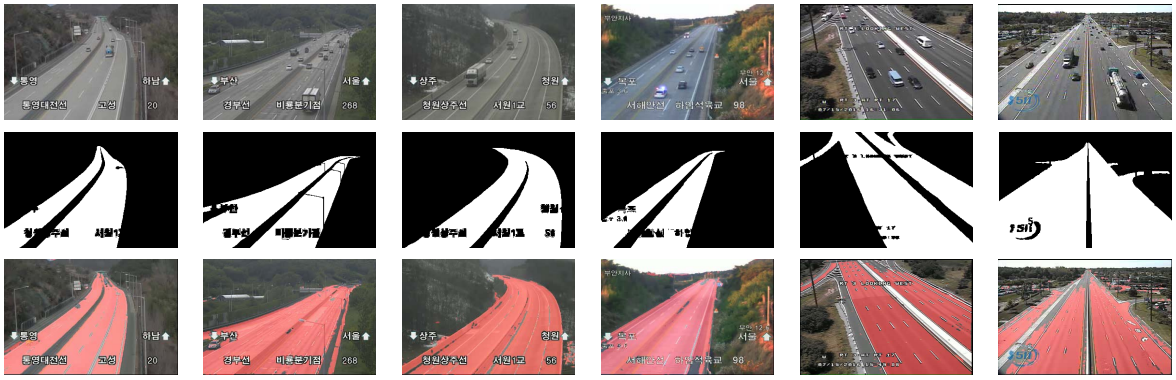
III. Experiments

In this section, the performance of the proposed method is evaluated on different videos with various illumination and weather conditions, resolution, and frame-rate values in order to ensure the diversity of the tested data. The used dataset, provided by New Jersey Department of Transportation (NJDOT), contains 84 real traffic surveillance videos with various illumination conditions, road shapes, resolutions, viewing angles, and frame-rates. A sample frame of each videos is displayed at the first rows of Figures 7 and 8. The ground-truth mask representing the road region corresponding to each video is illustrated at the second rows of Figures 7 and 8 and the third rows present the resulting extracted road as a red mask on the background image of each video. The experiments were carried out using a DELL XPS 8900 PC with a 3.4 GHz processor and 16 GB RAM. The average speed was ~ 42.22 frames per second for videos of size 720×480 pixels, which shows the feasibility of the proposed method for real-time applications.

In order to evaluate the quantitative results, several evaluation metrics are utilized as follows:

$$\begin{cases} FPR = F_P/(F_P + T_N) \\ PRE = T_P/(T_P + F_P) \\ REC = T_P/(T_P + F_N) \\ ACC = (T_P + T_N)/(T_P + F_P + T_N + F_N) \\ F_1 = 2 \times (PRE \times REC)/(PRE + REC) \end{cases} \tag{10}$$

where T_P, F_P refer to the number of pixels correctly and incorrectly detected as part of the road region, and T_N and F_N are the number of pixels that are correctly and incorrectly detected as part of the non-road region, respectively. FPR, PRE, REC, ACC, and F_1 refer to false positive rate, precision, recall, accuracy, and F-measure respectively. The number of pixels classified as road and non-road are compared with the ground-truth data to calculate each measure. Figure 6 demonstrates the accuracy, F1 score, and false-positive rate charts for a number of traffic videos. An instant improvement in the detection results can be seen in the charts shown in Figure 6 which corresponds to the frame at which the first vehicle is observed in the video and a number of pixels corresponding to the location of the vehicle can be used as initial road samples.



(a) Video 1 (b) Video 2 (c) Video 3 (d) Video 4 (e) Video 5 (f) Video 6

Fig. 7: Road extraction results in regular traffic videos. The first row displays a sample frame of each video. The second row represents the ground-truth road region masks. The third row illustrates the extracted road region by the proposed method before applying the accumulative foreground mask.

(a) Video 7 (b) Video 8 (c) Video 9 (d) Video 10 (e) Video 11 (f) Video 12

Fig. 8: Road extraction results in traffic videos with challenging illumination conditions. The first row displays a sample frame of each video. The second row represents the ground-truth road region masks. The third row illustrates the extracted road region by the proposed method before applying the accumulative foreground mask.

TABLE I: The quantitative evaluation of the proposed method

Video #	1	2	3	4	5	6	7	8	9	10	11	12	**Average**
Precision	0.98	0.87	1	0.93	0.99	0.99	0.86	0.89	0.97	0.80	0.97	0.99	0.94
Recall	0.96	0.93	0.95	0.94	0.87	0.73	0.98	0.96	0.89	0.92	0.89	0.91	0.91
F-Score	0.97	0.90	0.97	0.93	0.92	0.84	0.92	0.92	0.93	0.86	0.93	0.95	0.93

Table I shows the quantitative performance of the road extraction method given 12 sample traffic videos. The precision values are higher than the recall values in most cases, which means that the entire roadway region is not always extracted due to under-segmentation. Some examples can be seen in Figures 7(e), 7(f), 8(a) and 8(e) This is usually caused by the perspective view and losing the tracking information at the far side of the road. Also, strong cast shadows and congested traffic can result in excluding some road pixels at the initial frames from the road map (e.g., Figure 8(b)). In some videos, the recall value is higher than the precision, which means there are more false-positive cases than false-negative. This is due to the overestimation or leak segmentation which is in turn caused by inconspicuous edges and lack of sufficient gradient information at the road boundaries. Another reason is the illumination effect which makes the non-road regions such as sky have similar values to the road pixels. Some examples of this can be seen in Figures 7(b), 7(d), 8(e) and 8(f). Here, we do not make any presumptions about the shape of the road in order for the approach to work on unstructured roads. Therefore, segmentation errors cannot be avoided by restrictions based on geometric models.

IV. Conclusion

Region of interest (RoI) determination is an essential pre-processing step in most image and video analytic applications. In case of traffic video analysis, the RoI usually refers to the road region where the objects of interest, i.e., vehicles, are located. In this paper, an adaptive statistical approach is proposed in order to extract the road region in real-time and automatically without the need of manual input. The proposed method can be applied on different videos with various resolution, frame-rate, illumination, and weather conditions. The road region extraction is performed by using color and temporal features and with no assumptions about high-level features such as the structure of the roadway, which makes the approach adaptive to various road shapes. The extracted road region can further be utilized as the RoI in video analytic tasks, such as anomaly detection, incident detection, recognition of hazardous driving behavior, speed estimation, and vehicle counting. The experimental results using real traffic video sequences provided by NJDOT demonstrate the feasibility and computational efficiency of the proposed method.

Acknowledgment

This paper is partially supported by the NSF grant 1647170.

References

[1] C.-K. Chang, J. Zhao, and L. Itti, "Deepvp: Deep learning for vanishing point detection on 1 million street view images," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.

[2] T. Kim, Y.-W. Tai, and S.-E. Yoon, "Pca based computation of illumination-invariant space for road detection," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 632–640.

[3] T. Rateke and A. von Wangenheim, "Road surface detection and differentiation considering surface damages," *arXiv preprint arXiv:2006.13377*, 2020.

[4] 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.

[5] M. A. Helala, F. Z. Qureshi, and K. Q. Pu, "Automatic parsing of lane and road boundaries in challenging traffic scenes," *Journal of electronic imaging*, vol. 24, no. 5, p. 053020, 2015.

[6] M. Santos, M. Linder, L. Schnitman, U. Nunes, and L. Oliveira, "Learning to segment roads for traffic analysis in urban images," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 527–532.

[7] Q.-J. Kong, L. Zhou, G. Xiong, and F. Zhu, "Automatic road detection for highway surveillance using frequency-domain information," in *Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics*. IEEE, 2013, pp. 24–28.

[8] J.-S. Lee and T.-H. Park, "Fast road detection by cnn-based camera-lidar fusion and spherical coordinate transformation," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[9] L. Caltagirone, L. Svensson, M. Wahde, and M. Sanfridson, "Lidar-camera co-training for semi-supervised road detection," *arXiv preprint arXiv:1911.12597*, 2019.

[10] Z. Chen, J. Zhang, and D. Tao, "Progressive lidar adaptation for road detection," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 693–702, 2019.

[11] J. W. Perng, Y. W. Hsu, Y. Z. Yang, C. Y. Chen, and T. K. Yin, "Development of an embedded road boundary detection system based on deep learning," *Image and Vision Computing*, p. 103935, 2020.

[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] N. Y. Q. Abderrahim, S. Abderrahim, and A. Rida, "Road segmentation using u-net architecture," in *2020 IEEE International conference of Moroccan Geomatics (Morgeo)*. IEEE, 2020, pp. 1–4.

[14] H. Shi and C. Liu, "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.

[15] ——, "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.

[16] J. M. Á. Alvarez and A. M. Ĺopez, "Road detection based on illuminant invariance," *IEEE transactions on intelligent transportation systems*, vol. 12, no. 1, pp. 184–193, 2010.

[17] 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.

[18] 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.

[19] F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Computer Vision and Image Understanding*, vol. 93, no. 2, pp. 206–220, 2004.

[20] 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.

[21] 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.

[22] 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.

[23] Y. Yuan, Z. Jiang, and Q. Wang, "Video-based road detection via online structural learning," *Neurocomputing*, vol. 168, pp. 336–347, 2015.

[24] L. Nguyen, S. L. Phung, and A. Bouzerdoum, "Enhanced pixel-wise voting for image vanishing point detection in road scenes," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1852–1856.

[25] ——, "Efficient vanishing point estimation for unstructured road scenes," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2016, pp. 1–6.

[26] J. Duan and M. Viktor, "Real time road edges detection and road signs recognition," in *2015 International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, 2015, pp. 107–112.

[27] T. Y. Goh, S. N. Basah, H. Yazid, M. J. A. Safar, and F. S. A. Saad, "Performance analysis of image thresholding: Otsu technique," *Measurement*, vol. 114, pp. 298–307, 2018.